

UNIVERSITÀ DI PISA

---

Dipartimento di Informatica  
Corso di Laurea Magistrale in Informatica per l'Economia e per l'Azienda  
(Business Informatics)



TESI DI LAUREA

PROGETTAZIONE E REALIZZAZIONE DI UN  
DATA WAREHOUSE PER IL CONTROLLO DI QUALITÀ,  
PER UN'AZIENDA FARMACEUTICA

Relatore:  
Chiar.mo Prof. Franco TURINI

Candidato:  
Alfredo COPPOLA

Tutor aziendale:  
Dott. Marco LUCHINI

---

Anno Accademico 2015-2016



*Alla mia famiglia,  
perchè senza loro questa tesi  
non sarebbe mai stata scritta.*

Alfredo Coppola



# Riassunto

L'obiettivo dell'attività, svolta presso l'azienda SDG Group (*Strategic Decision Governance*), è la realizzazione di una piattaforma *software* in cui saranno disponibili informazioni rilevanti relative al controllo di qualità effettuato sui lotti di insulina.

La soluzione è stata progettata per essere utilizzata da differenti categorie di utenti *business*, per l'analisi dei *trend* di parametri specifici per il supporto al processo decisionale.

Il lavoro è introdotto con una presentazione della tematica, descrivendo il processo e gli strumenti utili per il controllo di qualità in un'azienda farmaceutica.

Successivamente sono descritti gli strumenti utilizzati, le attività di progettazione e implementazione del *data warehouse*, soffermandosi sul processo di elaborazione dei dati in ingresso e sulla rappresentazione grafica delle informazioni offerta agli utenti.



# Indice

<b>1</b>	<b>INTRODUZIONE</b>	<b>1</b>
1.1	Presentazione del problema . . . . .	1
1.2	Rassegna della letteratura . . . . .	1
1.3	Contenuto della tesi . . . . .	2
<b>2</b>	<b>CASO DI STUDIO</b>	<b>3</b>
2.1	Il controllo di qualità . . . . .	3
2.1.1	Il controllo statistico di qualità . . . . .	4
2.1.2	Il controllo di qualità nel campo farmaceutico . . . . .	4
2.1.3	Alcuni strumenti di analisi per il controllo di qualità . . . . .	6
2.2	L'azienda sviluppatrice: SDG Group . . . . .	12
2.3	L'azienda cliente: multinazionale farmaceutica . . . . .	12
2.3.1	Il processo produttivo . . . . .	13
<b>3</b>	<b>ARCHITETTURA DI DATA WAREHOUSE</b>	<b>17</b>
3.1	Il processo di <i>data warehousing</i> . . . . .	17
3.2	Architettura <i>global manufacturing data mart</i> . . . . .	18
3.2.1	<i>Source area</i> . . . . .	20
3.2.2	<i>Staging area</i> . . . . .	20
3.2.3	<i>Core area</i> . . . . .	21
3.2.4	<i>Extension area</i> . . . . .	22
3.2.5	Nomenclatura . . . . .	22
<b>4</b>	<b>ANALISI DEI REQUISITI E PROGETTAZIONE CONCETTUALE INIZIALE</b>	<b>25</b>
4.1	Analisi dei requisiti . . . . .	25
4.1.1	Specificazione dei requisiti . . . . .	26
4.2	Progettazione concettuale iniziale . . . . .	30
<b>5</b>	<b>PROGETTAZIONE CONCETTUALE FINALE E LOGICA</b>	<b>31</b>
5.1	Presentazione delle fonti di dato disponibili . . . . .	31
5.1.1	Fonte <i>IPC Book</i> . . . . .	32
5.1.2	Fonte <i>PMX</i> . . . . .	33
5.2	Progettazione concettuale candidata e finale . . . . .	33
5.3	Progettazione logica del <i>data mart</i> . . . . .	34
<b>6</b>	<b>AMBIENTE DI SVILUPPO</b>	<b>37</b>
6.1	Aspetti generali . . . . .	37
6.2	<i>Database Oracle</i> . . . . .	38
6.3	<i>Oracle SQL Developer</i> . . . . .	38
6.4	<i>Informatica PowerCenter</i> . . . . .	39
6.4.1	Architettura . . . . .	39

6.4.2	<i>Repository manager</i>	40
6.4.3	<i>Designer</i>	41
6.4.4	<i>Workflow manager</i>	45
6.4.5	<i>Workflow monitor</i>	47
6.5	Microsoft SQL Server Reporting Services Report Builder	48
<b>7</b>	<b>ESTRAZIONE, TRASFORMAZIONE E CARICAMENTO</b>	<b>51</b>
7.1	Integrazione dell'area di <i>staging</i> e del <i>core</i>	52
7.1.1	Integrazione dell'area di <i>staging</i>	53
7.1.2	Integrazione dell'area del <i>core</i>	56
7.2	Integrazione dell'area dell' <i>extension</i>	58
7.3	Schedulazione dei <i>work flow</i>	60
<b>8</b>	<b>INTERFACCIA UTENTE E REPORTISTICA</b>	<b>61</b>
8.1	Fasi principali per la realizzazione	61
8.2	Esempi d'uso	64
	<b>CONCLUSIONI</b>	<b>71</b>
<b>A</b>	<b>Estrazione, trasformazione e caricamento</b>	<b>73</b>
<b>B</b>	<b>File dei parametri</b>	<b>83</b>
<b>C</b>	<b>Schedulazione dei <i>work flow</i></b>	<b>85</b>
<b>D</b>	<b>Interfaccia utente e reportistica</b>	<b>87</b>
	<b>Bibliografia</b>	<b>95</b>



# Elenco delle figure

2.1	La carta di controllo . . . . .	7
2.2	Carta di controllo di un processo fuori controllo statistico . . . . .	7
2.3	1° caso di carta di controllo di un processo fuori controllo . . . . .	8
2.4	2° caso di carta di controllo di un processo fuori controllo . . . . .	8
2.5	3° caso di carta di controllo di un processo fuori controllo . . . . .	8
2.6	Test delle zone . . . . .	9
2.7	Istogramma di frequenze . . . . .	12
2.8	Logo di SDG Group . . . . .	12
2.9	Cartuccia che conterrà l'insulina . . . . .	13
2.10	Flusso dell'attività di <i>filling</i> . . . . .	14
2.11	<i>Automatic inspection visual machine</i> . . . . .	15
3.1	<i>Data warehousing</i> con architettura a tre livelli . . . . .	18
3.2	Architettura <i>global manufacturing data mart</i> . . . . .	19
4.1	Modello concettuale iniziale . . . . .	30
5.1	Schema logico del <i>data mart</i> analisi qualità . . . . .	34
6.1	Architettura di Informatica PowerCenter . . . . .	39
6.2	Finestra principale del <i>repository manager</i> . . . . .	40
6.3	Finestra principale del <i>repository manager</i> con le proprietà dell'oggetto selezionato . . . . .	41
6.4	Finestra principale del <i>designer</i> . . . . .	41
6.5	Esempio di <i>mapping</i> creato con il <i>designer</i> . . . . .	42
6.6	Finestra principale del <i>target designer</i> . . . . .	43
6.7	Concatenazione di due campi di tipo <i>string</i> . . . . .	44
6.8	Oggetto <i>expression</i> del <i>designer</i> . . . . .	45
6.9	Finestra principale del <i>workflow manager</i> . . . . .	46
6.10	Schedulatore del <i>workflow manager</i> . . . . .	47
6.11	Finestra principale del <i>workflow monitor</i> . . . . .	47
6.12	Finestra principale di Microsoft SQL Server Reporting Services Report Builder . . . . .	48
7.1	<i>Work flow</i> per l'integrazione dell'area di <i>staging</i> e del <i>core</i> . . . . .	52
7.2	<i>Worklet</i> w_ipc_gmdm_stg_FT contenuta nel <i>work flow</i> wf_ipc_gmdm . . . . .	53
7.3	<i>Mapping</i> m_STG_LOAD associato alla <i>session</i> s_TBSTFT_IPC_STATS_VIEW . . . . .	54
7.4	Oggetto sp_LOAD_STG per eseguire la <i>store procedure</i> LOAD_STG . . . . .	54
7.5	Restrizione sulla tabella di configurazione TBSTCNF_LOAD_STG . . . . .	55
7.6	Configurazione di esecuzioni concorrenti . . . . .	55
7.7	<i>Worklet</i> w_ipc_gmdm_core_FT contenuta nel <i>work flow</i> wf_ipc_gmdm . . . . .	56
7.8	<i>Mapping</i> per l'integrazione dell'area del <i>core</i> . . . . .	56
7.9	Oggetto <i>expression</i> definito nel <i>mapping</i> per l'integrazione dell'area del <i>core</i> . . . . .	57

7.10	Oggetto <i>expression</i> per la strategia di aggiornamento della tabella <i>target</i> . . . . .	57
7.11	Oggetto <i>update strategy</i> per la strategia di aggiornamento della tabella <i>target</i> . . .	58
7.12	Chiamata alla procedura di <i>delete</i> per la tabella del <i>core</i> . . . . .	58
7.13	<i>Work flow</i> per l'integrazione dell'area dell' <i>extension</i> . . . . .	58
7.14	Worklet <i>w_gmdm_ext_GFCC</i> contenuta nel <i>work flow</i> <i>wf_filling_control_chart</i> . .	59
7.15	<i>Mapping</i> per l'integrazione dell'area dell' <i>extension</i> . . . . .	59
7.16	Oggetto <i>expression</i> definito nel <i>mapping</i> per l'integrazione dell'area dell' <i>extension</i>	60
8.1	Definizione della prima carta di controllo . . . . .	62
8.2	Definizione della seconda carta di controllo . . . . .	62
8.3	Definizione dell'istogramma . . . . .	62
8.4	Condizione per il colore per l'istogramma . . . . .	63
8.5	Menù a tendina con la lista dei possibili <i>batch</i> . . . . .	63
8.6	Interfaccia <i>web</i> per le <i>global filling control charts</i> . . . . .	64
8.7	Finestra principale delle <i>global filling control charts</i> . . . . .	64
8.8	Esempio di carta di controllo per parametro <i>plunger depth</i> per <i>batch</i> di tipo Humalog Solution . . . . .	65
8.9	Esempio di carta di controllo con i valori di LSL e USL per parametro <i>plunger depth</i> per <i>batch</i> di tipo Humalog Solution . . . . .	65
8.10	Esempio di istogramma con i valori di $C_{pk}$ per parametro <i>plunger depth</i> per <i>batch</i> di tipo Humalog Solution . . . . .	66
8.11	Menù superiore . . . . .	66
8.12	Esempio di carta di controllo per parametro <i>total lenght</i> per <i>batch</i> di tipo Humalog Solution . . . . .	66
8.13	Esempio di carta di controllo con i valori di LSL e USL per parametro <i>total lenght</i> per <i>batch</i> di tipo Humalog Solution . . . . .	67
8.14	Esempio di istogramma con i valori di $C_{pk}$ per parametro <i>total lenght</i> per <i>batch</i> di tipo Humalog Solution . . . . .	67
8.15	Esempio di carta di controllo per parametro <i>plunger depth</i> per <i>sample</i> di tipo Humalog Solution . . . . .	68
8.16	Esempio di carta di controllo con i valori di LSL e USL per parametro <i>plunger depth</i> per <i>sample</i> di tipo Humalog Solution . . . . .	68
8.17	Esempio di istogramma con i valori di $C_{pk}$ per parametro <i>plunger depth</i> per <i>sample</i> di tipo Humalog Solution . . . . .	69

# Elenco delle tabelle

3.1	Principali sistemi sorgenti del GMDM . . . . .	20
4.1	Processo analisi qualità . . . . .	27
4.2	Fatto analisi qualità . . . . .	28
4.3	Dimensioni . . . . .	28
4.4	<i>Batch</i> . . . . .	28
4.5	<i>Phase</i> . . . . .	29
4.6	<i>Material</i> . . . . .	29
4.7	<i>Parameter</i> . . . . .	29
4.8	Gerarchie dimensionali . . . . .	29
4.9	Trattamento delle modifiche . . . . .	30
4.10	Misura . . . . .	30
5.1	Fonte IPC <i>Book</i> . . . . .	33
5.2	Tabella <i>data mart</i> della soluzione <i>global filling control charts</i> . . . . .	35



# Listings

A.1	Configurazione della tabella TBSTCNF_LOAD_STG . . . . .	73
A.2	SQL definito nel <i>source qualifier</i> del <i>mapping</i> per l'integrazione dell'area del <i>core</i> .	74
A.3	SQL definito nel <i>source qualifier</i> del <i>mapping</i> per l'integrazione dell'area dell' <i>extension</i>	75
A.4	Procedura di <i>delete</i> per la tabella del <i>core</i> . . . . .	78
A.5	Procedura di <i>delete</i> per la tabella dell' <i>extension</i> . . . . .	80
B.1	<i>File</i> dei parametri per il <i>work flow</i> wf_ipc_gmdm . . . . .	83
B.2	<i>File</i> dei parametri per il <i>work flow</i> wf_filling_control_chart . . . . .	84
C.1	Comandi <i>unix</i> per la schedulazione dei <i>work flow</i> . . . . .	85
D.1	<i>Data set</i> DataSetCCData . . . . .	87
D.2	<i>Data set</i> DataSetBatch . . . . .	94



# Capitolo 1

## INTRODUZIONE

### 1.1 Presentazione del problema

L'innovazione rappresenta l'elemento trainante di ogni settore tecnologicamente avanzato, ma in campo farmaceutico acquisisce una importanza particolare poiché riguarda la salute e la qualità della vita delle persone. Nasce la necessità di rispondere ai bisogni della società, anticipare le risposte e individuare soluzioni terapeutiche efficaci e sicure, con l'obiettivo di migliorare le condizioni di vita dei pazienti. Questo rappresenta una grande conquista per i pazienti che possono sperare in passi avanti consistenti nella cura delle patologie in tempi relativamente rapidi, ma rappresenta una sfida per le aziende alle prese con una competizione sempre più dura. Per questo motivo la tendenza delle principali case è quella di delineare accuratamente non solo la strategia, ma anche l'ambito in cui concentrare i propri sforzi. Ciò ha originato il bisogno, per le grandi aziende farmaceutiche, di utilizzare tutti gli strumenti a loro disposizione per operare efficientemente e primeggiare il mercato.

Nasce la necessità di creare valore servendosi dei dati e delle informazioni raccolte durante l'intero processo produttivo, servendosi di soluzioni tipiche della *business intelligence*. L'acquisizione, il mantenimento e la gestione della conoscenza diventa la variabile critica per il successo.

Il presente lavoro descrive un progetto svolto presso l'azienda SDG Group, volto alla realizzazione di un *data warehouse* per il monitoraggio di variabili rilevanti per il controllo di qualità sui lotti di insulina. Il controllo di qualità comprende l'insieme delle strutture organizzative, delle responsabilità, delle procedure, dei processi e delle risorse (umane e strumentali) messe in atto da un'azienda per realizzare e gestire la qualità, nel senso più generale del termine, dei propri prodotti e/o servizi.

Il sistema richiesto, che dovrà essere regolamentato dalle norme di buona fabbricazione, ha l'obiettivo di fornire a precise categorie di soggetti, una piattaforma unica in cui saranno disponibili informazioni rilevanti rappresentate tramite carte di controllo, consentendo di effettuare analisi e monitorare i risultati al fine di verificare la conformità del prodotto con le specifiche di approvazione garantendo che le caratteristiche del prodotto siano quelle richieste dalle normative.

### 1.2 Rassegna della letteratura

Nella stesura del presente documento sono state consultate numerose fonti.

Una parte sostanziosa del lavoro è basata sul contenuto di [Albano,2014], particolarmente adatto come guida alla realizzazione di progetti di *data warehousing* in quanto concetti teorici e pratici sull'argomento sono correlati da indicazioni sulla presentazione formale di quanto progettato.

Per l'introduzione di concetti relativi al controllo di qualità si è fatto riferimento a numerose opere come [Lanati,2010],[Lanzara,2012] e [Miolo Vitali,2000], ma anche alla normativa tecnica sulla qualità [UNI EN ISO 8402,1995].

Per la descrizione degli strumenti utilizzati, si è fatto riferimento a manuali tecnici, in particolare [Informatica,2013].

Le informazioni relative alle società sviluppatrice nel progetto sono estrapolate dal sito web ufficiale [SDG Group,2016].

### 1.3 Contenuto della tesi

Il presente documento ha l'obiettivo di presentare il lavoro svolto per il progetto di *data warehousing* finalizzato a supportare il controllo di qualità di una importante multinazionale farmaceutica, introducendo il lettore alla tematica, mostrando successivamente come quanto richiesto sia stato progettato e realizzato, evidenziando tecniche e tecnologie utilizzate e descrivendo l'interazione con l'utente finale.

In particolare il Capitolo 2 è dedicato alla presentazione del caso di studio e ha come obiettivo principale quello di introdurre ambito, soggetti ed altri aspetti relativi al progetto realizzato. Si definiscono pertanto inizialmente alcuni concetti chiave relativi alla tematica centrale del lavoro, cioè il controllo di qualità, l'attività di controllo in un'azienda farmaceutica e gli strumenti di analisi a supporto del controllo di qualità.

In seguito vengono presentate le organizzazioni interessate, nei ruoli di azienda committente e azienda sviluppatrice, e alcune caratteristiche generali del prodotto richiesto dal cliente.

Nel Capitolo 3 viene presentata l'architettura di *data warehouse* adottata dall'azienda cliente, al fine di comprendere meglio le scelte e le metodologie impiegate descritte nei capitoli successivi.

Partendo dai requisiti e dalle richieste espresse dal cliente, nel Capitolo 4, si presenta l'iniziale modello concettuale del *data mart* attraverso il formalismo del modello dimensionale dei fatti. Il modello usato prevede la descrizione del fatto, delle misure da utilizzare, delle dimensioni di analisi e delle gerarchie presenti in esse.

Il Capitolo 5 conclude la descrizione della progettazione concettuale finale e definisce la trasformazione di questa in un progetto logico, corrispondente a uno schema relazionale. Per arrivare a definire il progetto finale sono analizzate le fonti informative disponibili sui sistemi di origine dei dati.

Il Capitolo 6 è dedicato all'introduzione dei principali strumenti usati per lo sviluppo del sistema. Si descrivono quindi lo strumento usato per le operazioni di estrazione, trasformazione e caricamento dei dati ("Informatica PowerCenter") e la piattaforma dedicata alla realizzazione di report grafici ("Microsoft SQL Server Reporting Services Report Builder"). Dei vari strumenti si descrivono le funzionalità più importanti.

Nel Capitolo 7 si descrive il processo di ETL che include tutte le operazioni realizzate e le procedure implementate per integrare le sorgenti dato disponibili e caricare le informazioni richieste all'interno del *data mart* progettato. Vengono presentati i flussi realizzati e le tecnologie usate.

Infine nel Capitolo 8 viene descritta quella parte di sistema visibile agli utenti finali. Si descrivono inizialmente i meccanismi usati per la creazione della reportistica. Successivamente viene mostrato graficamente l'aspetto del *report* realizzato mostrando la personalizzazione grazie all'impostazione dei *prompt*. Il capitolo è corredato di esempi di *report* specifici per il supporto al processo decisionale.



## Capitolo 2

# CASO DI STUDIO

Il presente lavoro descrive un progetto svolto presso l'azienda SDG Group, volto alla realizzazione di un *data warehouse* per il monitoraggio di variabili rilevanti per il controllo di qualità sui lotti di insulina.

Il sistema richiesto, che dovrà essere regolamentato dalle norme di buona fabbricazione, ha l'obiettivo di fornire agli utenti del reparto qualità, una piattaforma unica per effettuare analisi di controllo al fine di rilasciare il prodotto finale conforme alle normative vigenti.

Nel presente capitolo, dopo aver inquadrato l'ambito di riferimento e le principali motivazioni del progetto, sono presentate l'azienda sviluppatrice e quella committente, della quale per motivi di *privacy* non è possibile indicarne il nome.

### 2.1 Il controllo di qualità

Realizzare un prodotto di qualità significa produrre rispettando certe specifiche e livelli di tolleranza prestabiliti, sulla base delle aspettative e preferenze dei clienti, che possono essere consumatori finali o nel caso di prodotti industriali le stesse aziende.

Si parla di esigenze espresse in quanto ci sono richieste che ciascuno di noi avanza esplicitamente ogni volta che deve fare un acquisto, ma ce ne sono altre, quelle implicite appunto, che vengono date per scontate e quindi non vengono neppure menzionate, ma che l'azienda deve comunque soddisfare [Miolo Vitali,2000].

Per produrre prodotti di alta qualità è necessario quindi conoscere le esigenze e i desideri dei clienti; questo è compito del dipartimento marketing di un'azienda. Successivamente sarà necessario tradurre tali esigenze in un progetto operativo, fino ad arrivare alla completa ingegnerizzazione del processo produttivo che condurrà, finalmente, alla realizzazione del prodotto fisico. Infine, l'organizzazione commerciale e distributiva provvederà alla consegna del prodotto al cliente.

Da quanto appena detto si capisce che, nell'ottica di quella che viene definita "qualità totale", produrre con alti livelli di qualità significa migliorare tutti i processi aziendali che contribuiscono alla produzione di prodotti o servizi, in vista della piena soddisfazione del cliente. Nello spirito della qualità totale, gestire per processi vuol dire curare al massimo grado i collegamenti fra le varie attività attraverso l'idea del "cliente interno". L'insieme delle attività che si svolgono nell'azienda viene considerato come un insieme di scambi fra clienti e fornitori interni.

Qualità totale significa non solo qualità esterna, per il cliente, ma anche qualità interna, per fare le cose bene la prima volta, per non sprecare risorse, per contrarre i tempi di risposta alle richieste del mercato, per creare un ambiente di lavoro in cui le persone siano motivate ad offrire il meglio delle proprie capacità [Miolo Vitali,2000].

La connotazione dinamica del mercato presuppone un continuo controllo e adeguamento del livello di qualità del prodotto. Diventa quindi necessario proporre un continuo miglioramento

della qualità, che richiede un investimento di tutta l'azienda nel suo complesso e che deve essere affrontato da tutte le funzioni aziendali in un'ottica integrata [Lanzara,2012].

In questo senso la strategia della qualità totale viene considerata come uno strumento per garantire la sopravvivenza e il successo dell'azienda nel lungo periodo. Le esigenze del mercato, infatti, non sono statiche e vengono ad essere, in qualche modo, correlate col livello stesso di qualità. Ad esempio, se la garanzia per le auto è di 3 anni per quasi tutte le marche, il mercato accetterà questo valore come uno standard di qualità. Se un produttore porta la garanzia a 5 anni lo standard tenderà ad alzarsi .

### 2.1.1 Il controllo statistico di qualità

In tale quadro di riferimento troverà una sua giustificazione l'applicazione di metodologie dirette al "controllo statistico del processo" in quanto sono esse a rappresentare un primo valido strumento a sostegno dell'attività di *decision making* ad ogni livello organizzativo per il raggiungimento della qualità. Come universalmente riconosciuto, il controllo statistico del processo (d'ora in poi SPC) può essere definito come una "metodologia che, in riferimento ad una determinata attività, operazione, fase o processo caratterizzato da ripetitività, fa ricorso a tecniche statistiche al fine di definire, analizzare e verificare le condizioni che determinano la variabilità dell'oggetto di analisi". In modo più sintetico, rifacendoci alla definizione fornita da Joseph M. Juran potremmo definire l'SPC come "l'applicazione di tecniche statistiche per comprendere ed analizzare le variabilità di un processo".

Gli studi sull'SPC non sono certo temi nuovi alla qualità. Già nel 1924, infatti, il dott. W.A. Shewart iniziò a sviluppare un approccio statistico al controllo di qualità, rilevando che il concetto di variabilità riferito ai fenomeni naturali era ugualmente adeguato all'analisi e alla descrizione dei processi produttivi. Con il contributo della statistica inferenziale e della statistica descrittiva, arrivò allora alla descrizione sintetica di fenomeni più ampi da impiegare come modelli di supporto alle attività di *problem solving*. Nacquero così le "carte di controllo", uno degli strumenti statistici più impiegati nell'analisi dei processi produttivi. Da allora i passi avanti compiuti sul tema sono stati molti. Primo fra tutti il riconoscimento circa la validità di questi strumenti e un loro più vasto impiego.

Il controllo statistico della qualità ha cessato di essere semplicemente un supporto al cosiddetto "*scientific management*" per divenire strumento diffuso da collocare all'interno di un vero e proprio approccio di gestione/organizzazione. In quest'ambito, le metodologie SPC seppur a livelli differenti di approfondimento sono divenute patrimonio aziendale comune e condiviso a tutti i livelli. Si è, in pratica, andato diffondendo all'interno dell'organizzazione un orientamento finalizzato a coniugare l'approccio tradizionale ai problemi con un approccio fondato sullo *statistical thinking* come atteggiamento culturale. Va sottolineato, tuttavia, come ciò valga non solo per ruoli tecnici piuttosto che manageriali, ma anche e soprattutto per coloro che, in quanto operatori, possono incidere direttamente sul proprio processo attraverso un'analisi che si configura come un vero e proprio *learning by doing*.

### 2.1.2 Il controllo di qualità nel campo farmaceutico

L'attività di controllo in una azienda farmaceutica, è un processo che segue tutta la vita del farmaco, dal suo sviluppo fino al suo arrivo in farmacia. I prodotti farmaceutici devono essere preparati in modo tale che non comportino rischi per i consumatori finali e che la loro qualità risponda a tutti i requisiti necessari ed indispensabili per l'uso a cui sono destinati. Per il raggiungimento di tale obiettivo è richiesto l'impegno di tutte le persone coinvolte a qualsiasi livello e un sistema di controllo qualità, chiaramente definito e correttamente applicato, coerente con le "norme di buona fabbricazione" (*good manufacturing practise*).

Il controllo di qualità inteso come le "tecniche e le attività a carattere operativo messe in atto per soddisfare i requisiti per la qualità" [UNI EN ISO 8402,1995], assicura che tutti i prodotti siano manifatturati e studiati secondo degli standard e delle regole molto severi.

Si esamina la produzione, vengono monitorati, e validati i processi manifatturieri e i sistemi informatici. Gli addetti del controllo qualità sono responsabili di assicurare la conformità delle materie prime, dei semilavorati e del prodotto finito agli standard qualitativi. Compilano la documentazione riguardante le procedure di laboratorio e riportano ogni anomalia. Revisionano e aggiornano anche le procedure standard operative e si possono occupare di progetti di tipo analitico e *problem solving* per tutta la strumentazione.

Un farmaco di qualità deve essere:

- **Efficace**, ossia deve svolgere l'azione terapeutica prevista.
- **Sicuro**, ovvero non deve mettere a rischio la salute del paziente, rischio che si può verificare nel caso in cui le materie prime utilizzate per la fabbricazione non abbiano la qualità prevista o il processo di fabbricazione non sia stato eseguito in maniera corretta.
- **Controllato**, ossia deve avere alle spalle un sistema di qualità, cioè un modo di lavorare corretto e standardizzato da parte di tutti coloro che partecipano alla fabbricazione del farmaco, che ne garantisca la costante corrispondenza alle specifiche approvate dalle autorità sanitarie.

Solo grazie ad un lavoro costante di tutte le funzioni coinvolte nei processi di produzione e controllo, è possibile garantire che il farmaco che verrà immesso sul mercato risponderà a tutti i requisiti previsti.

Nel campo farmaceutico, cresce sempre di più l'uso di sistemi informatici al fine di intercettare nel modo più rapido possibile un qualunque evento che pregiudichi l'accuratezza (esattezza e precisione) del metodo analitico correntemente utilizzato.

Obiettivo dell'impresa, è quello di avere strumenti adeguati per migliorare il livello dei prodotti/servizi offerti/erogati attraverso l'eliminazione di errori, difformità che causano ripetizioni di lavoro, controlli inutili e quindi rallentamenti nei cicli di lavorazione. Questo è realizzabile grazie l'impiego di dati significativi, in quanto analizzabili, che consentano di determinare e interpretare performance e cause che determinano "cambiamenti indesiderati" rispetto al normale funzionamento del processo in analisi.

### Le norme di buona fabbricazione

Le norme di buona fabbricazione (d'ora in poi GMP), sono costituite da un insieme di regole e linee guida relative alla qualità di un prodotto farmaceutico o di sostanza farmacologicamente attive. Lo scopo per il quale tali linee guida sono state create è assicurare che un farmaco sia prodotto, analizzato e rilasciato in un regime di qualità controllata e certificata (nonché verificabile), in modo da assicurare il livello qualitativo; in questo modo, è ridotto al minimo il pericolo che vi siano rischi non previsti e non controllati per il paziente [Lanati,2010].

Al fine di ottenere una certificazione GMP, l'azienda deve dimostrare che essa stessa provvede alla corretta manutenzione degli impianti e delle attrezzature e che segue le procedure approvate per tutte le operazioni, dall'ordine e dalla ricezione di materie prime alla produzione, confezionamento, stoccaggio, spedizione e consegna.

Il regolamento GMP disciplina settori come la qualifica del personale, igiene, pulizia, verifica delle apparecchiature, convalida dei processi e la gestione dei reclami. Le GMP richiedono documentazione approfondita per ogni aspetto dello sviluppo, della sperimentazione, della produzione e della logistica per consentire la tracciabilità e l'eventuale richiamo di un prodotto dal mercato nel caso in cui dovessero sorgere dei problemi.

"Un'azienda farmaceutica con una cultura della qualità altamente sviluppata spende, in media, 350 milioni di dollari in meno per correggere gli errori rispetto a un'azienda farmaceutica con una cultura della qualità meno sviluppata."

Creare una cultura della qualità, Harvard Business Review, aprile 2014

### 2.1.3 Alcuni strumenti di analisi per il controllo di qualità

#### Le carte di controllo

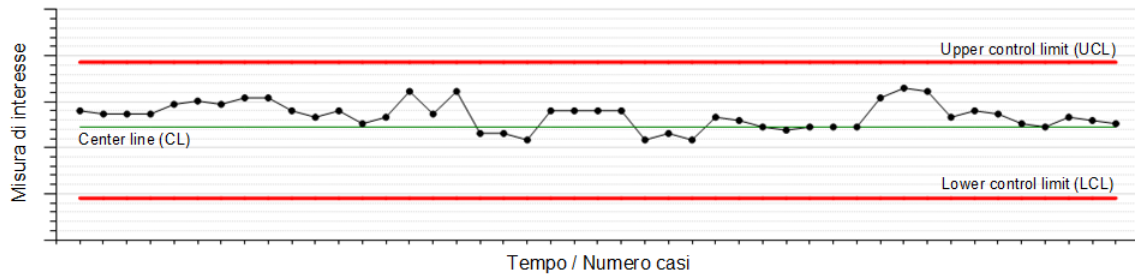
Ogni processo produttivo è caratterizzato da una certa variabilità naturale, questa variabilità è presente anche se il processo è ben progettato e controllato ed è dovuta all'azione congiunta di molte piccole cause e generalmente non è addebitabile a singoli fattori controllabili: usualmente in queste condizioni tale variabilità è piccola. Quando un processo produttivo è caratterizzato solo da una variabilità naturale, si può affermare che il processo opera soggetto ad un sistema di cause accidentali o comuni.

Nella terminologia del controllo statistico di processo, un processo che opera soggetto solo ad un sistema di cause accidentali è in uno "stato di controllo statistico". Altre fonti di variabilità, dovute a fattori ben individuabili e controllabili, possono intervenire nel processo produttivo alterando ed aumentando la variabilità "naturale" fino a valori non accettabili per gli standard di qualità. In questo caso si può affermare che il processo opera soggetto ad un insieme di cause sistematiche o speciali. Un processo che opera in presenza di cause sistematiche è in uno "stato di fuori controllo statistico".

La carta di controllo, supportando l'analista nel riconoscimento della causa di variazione, consente di individuare un processo fuori controllo consentendo di evitare due errori tipici. Il primo consiste nell'interpretare una causa comune di variazione come una causa straordinaria; in questo caso si potrebbe esercitare un'azione correttiva eccessiva che può a sua volta aumentare la variabilità del processo. Il secondo è l'errore speculare, e si commette quando una variazione straordinaria viene trattata come una variazione comune. In questo caso, si rischia di non intervenire prontamente e adeguatamente per "correggere" il sistema.

Le carte di controllo rappresentano uno dei metodi più utilizzati per il controllo statistico di produzione. Messe a punto negli anni '30 da Walter Shewthard, il loro utilizzo si è rapidamente diffuso negli Stati Uniti e poi in Giappone già prima della seconda guerra mondiale. Negli anni successivi alla guerra la loro utilità è stata messa in discussione per il fatto di non fornire alcuno strumento per la risoluzione dei problemi. Tuttavia ancora oggi le carte di controllo sono lo strumento principe nel controllo statistico di processo.

Le carte di controllo sono essenzialmente rappresentazioni grafiche di un processo nel tempo che, basandosi su teorie statistiche, rimangono di facile interpretazione e utilizzo anche per utenti meno esperti. In letteratura sono state proposte numerose carte di controllo che differiscono tra loro sia per le modalità di costruzione che per l'obiettivo che viene perseguito con il loro utilizzo. L'elemento che le accomuna è l'interpretazione in termini di serie storiche della successione di osservazioni condotte sull'output del processo [Coccarda,2011] (Fig. 2.1).



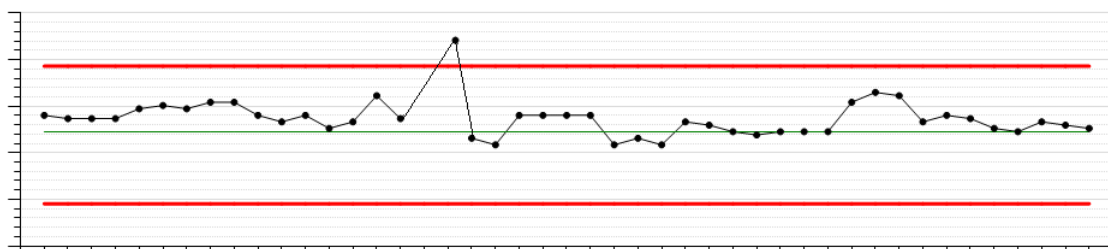
**Figura 2.1:** La carta di controllo

Nella carta è presente una linea centrale, CL (*central line*), che rappresenta il valore medio caratteristica di qualità in genere corrispondente al valore desiderato nell'ipotesi di controllo del processo. Altre due linee orizzontali identificano i limiti di controllo: UCL (*upper control limit*) il limite di controllo superiore e LCL (*lower control limit*) il limite di controllo inferiore.

Una volta definiti i limiti di controllo, plottando i dati all'interno del grafico, la carta consente di individuare eventuali andamenti sistematici (*pattern*) dei valori che rappresentano il processo nel tempo e di stabilire se ciascun punto cade all'interno o all'esterno dei limiti imposti. In questo modo si individua immediatamente un processo fuori controllo.

Nella pratica raramente vengono considerate singole osservazioni, il più delle volte i valori riportati nelle carte di controllo rappresentano a loro volta il risultato di una stima campionaria (pertanto il singolo valore riportato nella carta di controllo non rappresenta una singola misura, ma, ad esempio, la media di una serie di misure eseguite campionando vari elementi dallo stesso lotto). I vari campioni utilizzati (che possono essere più o meno grandi in funzione della produzione presa in esame) vengono chiamati sottogruppi e si presume che al loro interno presentino la sola variabilità casuale, mentre le cause speciali di variazione (se presenti) possono determinare solamente la variabilità fra diversi sottogruppi.

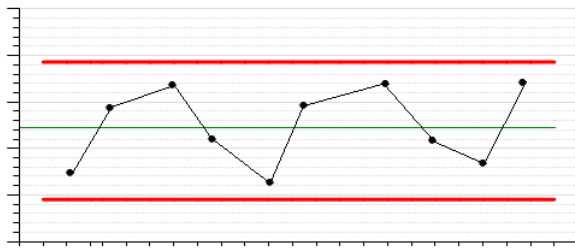
Il grafico di figura 2.2 mostra un tipico esempio di carta di controllo in cui si ha una misura fuori controllo (un punto fuoriesce dai limiti di controllo), questo è sintomatico di un processo fuori controllo statistico sulla cui variabilità intervengono cause speciali di variazione.



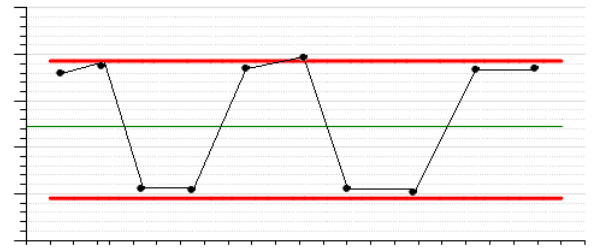
**Figura 2.2:** Carta di controllo di un processo fuori controllo statistico

La presenza di tutti i punti della carta all'interno dei limiti di controllo è solo una condizione necessaria ma non sufficiente per poter dire che un processo è in controllo. Si possono individuare sequenze temporali particolari che evidenziano che qualcosa di anomalo sta intervenendo nel processo.

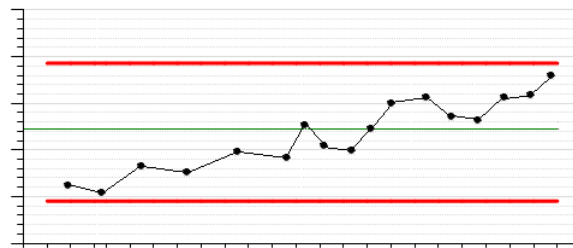
Può succedere che la maggioranza dei punti è collocata al di sotto del limite centrale, questo induce a pensare ad un'anomalia nel processo, dovuta ad esempio ad un operatore inesperto o ad un difetto di macchina. Situazioni analoghe (Fig. 2.3) possono presentarsi anche nel caso in cui si evidenzia un andamento ciclico dei dati dovuto sicuramente a delle cause ricorrenti (un ambiente climaticamente non controllato durante le stagioni estreme può essere un esempio di causa) o nel caso in cui i punti del tracciato tendano a cadere vicino ai due limiti di controllo con assenza di punti nel centro (si parla di mistura), creando un effetto come quello mostrato in figura 2.4 (questo evidenzia la presenza di due popolazioni distinte che possono essere, ad esempio, il risultato di due operatori che settano la macchina in modo diverso all'inizio dei turni).



**Figura 2.3:** 1° caso di carta di controllo di un processo fuori controllo



**Figura 2.4:** 2° caso di carta di controllo di un processo fuori controllo

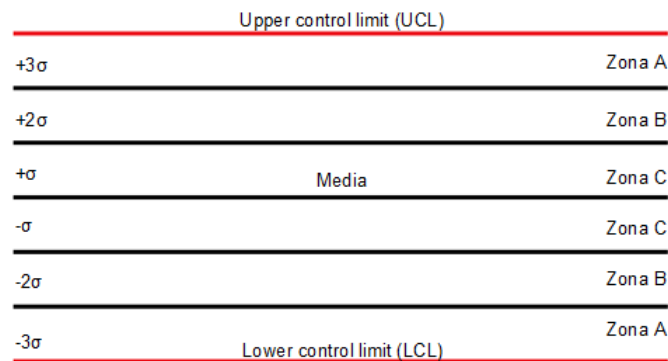


**Figura 2.5:** 3° caso di carta di controllo di un processo fuori controllo

Un caso ancora più evidente si verifica quando il tracciato dei valori segue un *trend* di crescita (o di decrescita) indicando una anomalia di processo dovuta ad esempio all'usura dell'utensile (si parla di tendenza), ad una diminuzione dell'abilità dell'operatore o ad un peggioramento dell'omogeneità del materiale (Fig. 2.5).

L'individuazione di eventuali anomalie nel tracciato di dati di una carta di controllo non è comunque compito semplice e necessita di esperienza e di una buona conoscenza del processo in esame.

Esistono però delle semplici regole pratiche a supporto del lavoro dell'analista, come il test delle zone (Fig. 2.6).



**Figura 2.6:** Test delle zone

Questo test è costruito in base al presupposto che la probabilità che una serie successiva di punti cada, per effetto del caso, in una data zona della carta di controllo sia bassa: ciò consente di concludere che siamo in presenza di una causa speciale e non di una causa comune di variazione.

Un processo è fuori controllo nei seguenti casi:

- 1) uno o più punti cadono fuori dei limiti di controllo;
- 2) nel caso in cui la carta di controllo sia divisa in zone, quando:
  - a) 2 punti, tra 3 consecutivi, sono nella zona A, dalla stessa parte rispetto al valor medio;
  - b) 4 punti, tra 5 consecutivi, sono dalla stessa parte rispetto al valor medio e in zona B;
  - c) 9 punti consecutivi sono dalla stessa parte rispetto al valor medio;
  - d) ci sono 6 punti consecutivi crescenti o decrescenti;
  - e) ci sono 14 punti consecutivi che si alternano su e giù;
  - f) ci sono 15 punti consecutivi in zona C (sopra e sotto il valor medio).

Una volta riconosciuto il fuori controllo del processo, compito dell'analista è di identificare le cause straordinarie di variazione. Se questi fattori determinano un peggioramento del prodotto, l'esperto deve pianificare strategie volte a eliminarli; se al contrario i fattori straordinari di variazione occorsi hanno portato a un miglioramento della qualità, il processo deve essere modificato in modo da incorporarli al suo interno. In questo modo, la causa straordinaria di variazione del vecchio processo diventa nel processo modificato una causa comune di variazione. I limiti di controllo andranno modificati (è necessario ricalcolare tutti gli simatori) quando sono state trovate e rimosse le cause di fuori controllo, fino a quando il processo non viene cambiato. In questo caso, nel nuovo calcolo vanno tenuti in considerazione solo i dati del nuovo processo.

Le carte di controllo comunemente utilizzate prendono il nome da Shewhart, il quale per primo utilizzò i dati a sua disposizione formulando diversi modelli grafici che si differenziano in base alle caratteristiche stesse dei dati e che sostanzialmente si dividono in due gruppi:

- per **variabili**, utilizzano delle misure quantitative;
- per **attributi**, utilizzano delle misure qualitative.

### La capacità di processo

Finora abbiamo visto risultati sulla stabilità e la consistenza del processo, ma nulla è stato detto circa l'accettabilità, è infatti possibile avere un processo stabile e consistente che però produce il 100% di prodotti inaccettabili.

La capacità del processo si riferisce all'uniformità. Gli studi relativi alla capacità del processo sono utilizzati per confrontare la variazione naturale dei dati individuali rispetto alle tolleranze ingegneristiche e indicano cosa il processo è capace di produrre e non cosa sta producendo realmente. Può essere utile avere un metodo quantitativo per esprimere la capacità del processo.

Per questo sono stati introdotti gli indici  $C_p$  e  $C_{pk}$ . Per calcolare questi indici è necessario che siano soddisfatte alcune ipotesi:

- il processo sia in controllo;
- l'istogramma dei dati sia gaussiano, vale a dire i valori tendono ad accentrarsi attorno alla loro media;
- siano noti i limiti di specifica.

Si è già detto come nell'attuale scenario economico è il cliente a stabilire la qualità di un servizio o di un prodotto. Il management di un'azienda deve prestare ascolto al cliente per poterne tradurre i bisogni e le aspettative in caratteristiche facilmente misurabili. Il management determina poi i limiti della specificazione di queste caratteristiche.

I limiti di specificazione rappresentano, dunque, le specificazioni tecniche che il management fissa in risposta ai bisogni e alle aspettative dei consumatori.

Il "limite di specificazione superiore" (*upper specification limit*, d'ora in poi USL) è il più grande dei valori che una caratteristica, oggetto di analisi, può assumere conformemente alle aspettative del consumatore. Il "limite di specificazione inferiore" (*lower specification limit*, d'ora in poi LSL) è il più piccolo dei valori che una caratteristica di interesse può assumere conformemente alle aspettative del consumatore.

Quindi, i limiti di specificazione non dipendono dalla popolazione ovvero non dipendono dal modello distributivo (e cioè dai suoi parametri) ma sono stabiliti all'esterno del processo.

La capacità del processo si riferisce, allora, alla capacità che lo stesso ha di soddisfare le richieste dei clienti.

Il primo indice di capacità usualmente impiegato nel controllo statistico di qualità è definito come:  $C_p = \frac{USL - LSL}{6\sigma}$ .

L'indice  $C_p$  confronta l'ampiezza dell'intervallo di conformità, cioè la tolleranza ingegneristica (numeratore) con la variabilità naturale del processo rappresentata dal valore  $6\sigma$ , detta anche "tolleranza naturale" (denominatore). A questo proposito è importante dire che, la grandezza  $6\sigma$  è considerata una misura della cosiddetta variabilità naturale del processo in stato di sotto controllo (99.7% se gaussiano).

La capacità del processo  $C_p$  è una misura della capacità del processo di costruire prodotti che soddisfano le specifiche e non tiene conto di dove si trova la media rispetto alle specifiche. Il processo si dice centrato se  $\mu = (USL + LSL)/2$  (cioè se la media cade nel mezzo dei limiti di specifica).  $USL - LSL$  rappresenta la tolleranza ingegneristica.

$C_p$  confronta la tolleranza ingegneristica con la variabilità naturale: se il processo è centrato e gaussiano,  $C_p > 1$ , significa che il processo è capace di produrre quasi il 100% di prodotti accettabili.

Le caratteristiche del processo in base al valore assunto da  $C_p$  sono:

- per  $C_p > 1$ : il processo è più che capace di produrre la quasi totalità dei prodotti entro le tolleranze;



- per  $C_p = 1$ : il processo è appena capace di produrre entro le tolleranze;
- per  $C_p < 1$ : il processo non è capace di produrre il 100% di risultati accettabili.

L'indice di capacità  $C_p$  non considera la posizione della media ( $\mu$ ) del processo rispetto ai limiti di specificazione. Di conseguenza può accadere che processi con il medesimo valore di  $C_p$  abbiano una frazione di elementi non conformi diversa. In altre parole, anche con valori  $C_p > 1$  si possono avere delle frazioni di elementi non conformi elevate, se la media del processo non è centrata rispetto ai limiti USL e LSL. La situazione può essere migliorata definendo una nuova misura di capacità che tenga conto della posizione di  $\mu$  rispetto a USL e LSL, in modo da fornire una relazione diretta tra l'indice e la frazione di non conforme. Questo indice, denominato  $C_{pk}$  è definito nel modo seguente:  $C_{pk} = \frac{\min\{\mu - LSL, USL - \mu\}}{3\sigma}$ .

$C_{pk}$  confronta la distanza media del processo dal limite di specifica più vicino, rapportato alla metà della dispersione del processo (spesso, chiamata variabilità 3-sigma).

L'indice  $C_{pk}$  è una misura della capacità "effettiva" del processo a differenza di  $C_p$  che misura la capacità "potenziale".

Le caratteristiche del processo in base al valore assunto da  $C_{pk}$  sono:

- per  $C_{pk} > 1$ : il processo è più che capace di produrre entro le tolleranze specificate;
- per  $C_{pk} = 1$ : il processo è appena capace di produrre entro le tolleranze specificate;
- per  $C_{pk} < 1$ : il processo non è capace di produrre la quasi totalità dei prodotti entro le tolleranze specificate;
- per  $C_{pk} < 0$ : la media del processo è all'esterno dei limiti di specifica.

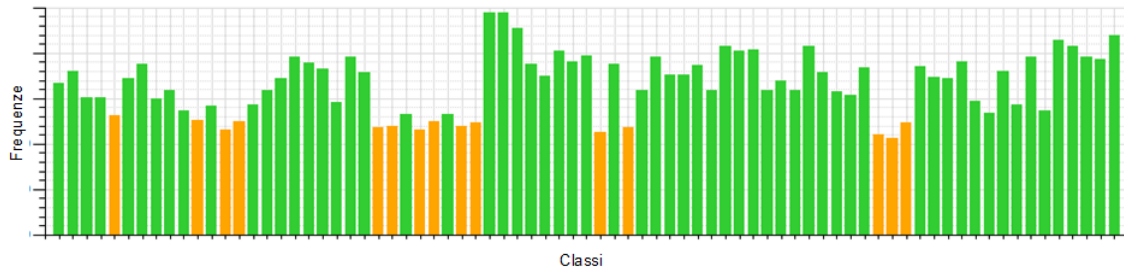
L'utilizzo di questa misura risulta spesso comodo se fatto in congiunzione con le carte di controllo, poiché in questo modo, oltre a non dover raccogliere appositamente i dati, la capacità del processo può essere analizzata indipendentemente dalla presenza di cause speciali di variazione (opportunamente individuate dalle carte) la cui influenza incide su variabilità e centratura del processo.

### L'istogramma di frequenze

Un semplice strumento impiegato per l'analisi di capacità di processo è l'istogramma di frequenze.

L'istogramma, appunto, è uno strumento grafico che consente di avere una visione completa e sintetica dei dati raccolti fornendo anche un indirizzo all'analisi delle cause. Può essere allora utile monitorare, attraverso un istogramma, la dispersione delle variabili più importanti e critiche del processo produttivo in modo da definire ipotesi e contromisure per la risoluzione del problema.

L'istogramma, che non è altro che un diagramma a colonne (Fig. 2.7), presenta in ordinata il numero di osservazioni in ciascuna classe e in ascissa le classi (il centro di ogni colonna coincide con il valore centrale della classe). Dove per classe si intende la dimensione di un intervallo di variabilità dei dati che si è preso come base per la rappresentazione dei dati stessi.



**Figura 2.7:** Istimogramma di frequenze

Considerare l'istogramma come uno strumento solo di rappresentazione e non di analisi è fortemente sbagliato. Dopo avere costruito l'istogramma occorre infatti trarne informazioni utili; spesso possiamo rilevare l'esistenza di problemi nel processo in esame a seconda dell'aspetto della distribuzione.

## 2.2 L'azienda sviluppatrice: SDG Group

SDG Group nasce nel 1991. Fin dal primo giorno, SDG è stata riconosciuta, dalla comunità di settore, come un fervente laboratorio di innovazione e sviluppo delle pratiche e metodologie di *planning and controlling*, così come un'emergente società di consulenza focalizzata sul *design* e l'implementazione di *decision support systems* [SDG Group, 2016].

Nel 1997 è stata delineata una nuova strategia di crescita con un chiaro focus sulle soluzioni di *business intelligence*. L'azienda sceglie di dedicarsi allo sviluppo di innovativi servizi di consulenza di *management intelligence* e *knowledge management*.

I servizi e le soluzioni offerte da SDG si sono progressivamente arricchiti negli anni con l'aggiunta di *social intelligence practices*, *big data architecture*, *business analytics* d'avanguardia e *digital transformation roadmaps*.

In quest'ottica SDG Group si impegna a diventare leader nell'offerta di metodi e strumenti di *business intelligence* grazie all'unione di specifiche capacità di *information technology* e consulenza gestionale, capace di soddisfare le necessità di ogni specifico settore industriale, aprendo la strada a nuove soluzioni per una migliore analisi dei dati.



**Figura 2.8:** Logo di SDG Group

## 2.3 L'azienda cliente: multinazionale farmaceutica

L'azienda committente è una multinazionale farmaceutica di origine statunitense, fondata sul finire dell'800. La società ha stabilimenti produttivi in oltre 10 paesi.

L'attività dell'affiliata italiana è iniziata circa 50 anni fa, per la produzione di farmaci da biotecnologia, destinate alla produzione di insulina da DNA ricombinante per i paesi europei ed extraeuropei. Lo stabilimento di produzione aveva il compito di manifatturare molti prodotti diversi in varie formulazioni e forme farmaceutiche, iniettabili e orali, sia per uso umano che veterinario, con un'alta complessità in termini di tecnologia farmaceutica per il mercato italiano.

Nel 2003, per far fronte alla domanda in continua crescita di insulina per il trattamento del diabete e alle necessità dell'ambiente italiano, viene deciso di specializzare lo stabilimento nella produzione di insulina ed analoghi da biotecnologia in cartucce e dispositivi di somministrazione pre-riempiti.

Il nuovo stabilimento, progettato sul principio dell'eccellenza, ha impianti all'avanguardia per prodotti sterili da biotecnologia e presenta un'architettura dei sistemi informatici integrata con l'automazione di processo e istruzioni di lavoro totalmente elettroniche.

L'azienda è orientata all'eccellenza per rendere la vita migliore alle persone di tutto il mondo. Per questo investe circa il 20% delle vendite nella scoperta e nello sviluppo di medicine innovative di alta qualità che aiutino le persone a vivere meglio e più a lungo.

### 2.3.1 Il processo produttivo

Il processo produttivo delle cartucce di insulina, nello stabilimento italiano dell'azienda cliente, prevede tre processi:

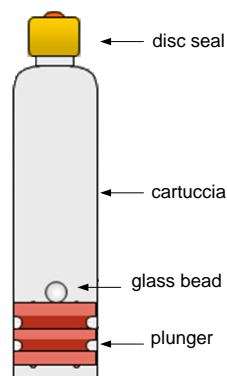
- **formulation:** è il processo in cui le diverse sostanze chimiche, tra cui il principio attivo dell'insulina, sono combinate per produrre la preparazione liquida;
- **filling:** è la fase di riempimento delle cartucce con il preparato realizzato nel processo precedente;
- **sorting:** si occupa dell'ispezione e del controllo qualità, per identificare eventuali anomalie e criticità nelle cartucce di insulina emerse nelle due fasi antecedenti.

Le cartucce che superano con successo il controllo della fase di *sorting* sono pronte per passare alla successiva fase, il *packaging*, che si occupa di assemblare e confezionare il prodotto finale.

Il sistema realizzato conterrà informazioni rilevanti analizzate nella sola fase di *filling*.

#### Attività di *filling*

Al fine di comprendere meglio questo elaborato, verranno presentati concetti chiave riguardanti l'attività di *filling*. Per la riservatezza dei metodi e dei processi di produzione, i concetti presentati saranno molto generali.



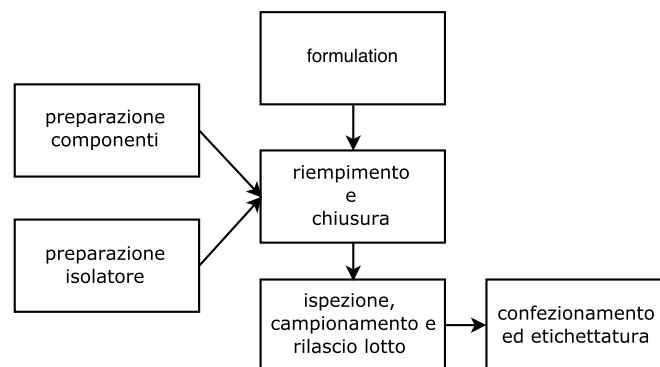
**Figura 2.9:** Cartuccia che conterrà l'insulina

La cartuccia che conterrà l'insulina è formata da 4 elementi fondamentali (Fig. 2.9):

- *disc seal*, la guarnizione di chiusura per garantire l'integrità e la protezione dell'insulina all'interno della cartuccia;
- *cartuccia*, contenitore in vetro pronto per essere riempito e inserito in dispositivi di iniezione;
- *glass bead*, perline di vetro pure per garantire la miscelazione;
- *plunger*, pistone per erogare l'insulina all'interno della cartuccia.

Le cartucce prodotte, possono essere utilizzate in specifici strumenti dosatori (*devices*) divisi in:

- penne monouso;
- penne multiuso (quindi ricaricabili).



**Figura 2.10:** Flusso dell'attività di *filling*

Il processo di riempimento, completamente automatico, segue la parte di *formulation* e precede l'attività di *sorting* (Fig. 2.10).

Il processo inizia con la preparazione dei componenti. Attraverso grosse lavatrici inox tutti i componenti che compongono una cartuccia, vengono lavati e sterilizzati. In questa fase iniziale avviene la siliconatura del *plunger*, per permettere lo scorrimento di esso all'interno di tutta la cartuccia.

Successivamente i componenti passano attraverso un tunnel di depirogenazione, dove le temperature del calore secco possono superare i 250°. La depirogenazione oltre a distruggere completamente tutti i microrganismi, permette la polimerizzazione del silicone creando una patina invisibile.

L'isolatore è una "scatola sterile chiusa" (che collega l'area di *formulation* con quella di *filling*) contenente la linea di riempimento, che va preparata, sterilizzata e mantenuta in questa condizione al fine di evitare ogni tipo di contaminazione.

La preparazione comprende:

- pulizia (riduzione carica batterica);
- assemblaggio linea;
- *set-up*;
- chiusura.

Tramite guide di trasporto, la cartuccia viene composta con il *plunger* e con aghi comandati da un sistema di vuoto vengono inserite le *glass bead*. A questo punto la cartuccia è pronta per essere riempita.

La fase di *filling* è stata studiata per:

- ottimizzare il tempo di riempimento;
- evitare formazione di schiuma;
- minimizzare la comparsa di bolle di aria;
- garantire alta precisione nella dosatura, indipendentemente da dimensioni dei materiali e dalle tolleranze di assemblaggio.

Dopo il riempimento, avviene la "tappatura" della cartuccia con il *disc seal*.

L'attività di controllo qualità nella fase di *filling* comprende:

- ispezione cartucce;
- stazione di scarto;
- stazione di campionamento.

L'ispezione della cartucce avviene tramite l'AVIM (*automatic visual inspection machine*) della Seidenader (Fig. 2.11), leader mondiale per la fabbricazione di macchinari per l'ispezione [Seidenader,2016].



**Figura 2.11:** *Automatic inspection visual machine*

Questa macchina che riesce a controllare fino a 600 cartucce al minuto, tramite 7 telecamere riesce a rilevare anomalie come:

- difetti di chiusura;
- crepe e graffi nella cartuccia;

- qualità della guarnizione di chiusura;
- rottura alla base o al corpo della cartuccia;
- perdite;

Dopo il controllo di ispezione, l'uscita della macchina prevede otto stazioni (quattro per le buone, una per i campioni, una per gli scarti e due per le dubbie che vanno re-ispezionate).

Le cartucce buone verranno sottoposte nuovamente a controlli di qualità più accurati e selettivi nella fase di *sorting*.

## Capitolo 3

# ARCHITETTURA DI *DATA* WAREHOUSE

Dopo una breve introduzione sulle caratteristiche e aspetti architettureali del processo di *data warehousing*, viene presentata l'architettura di *data warehouse* adottata dall'azienda cliente, al fine di comprendere meglio le scelte e le metodologie impiegate per la realizzazione del sistema denominato "*global filling control charts*".

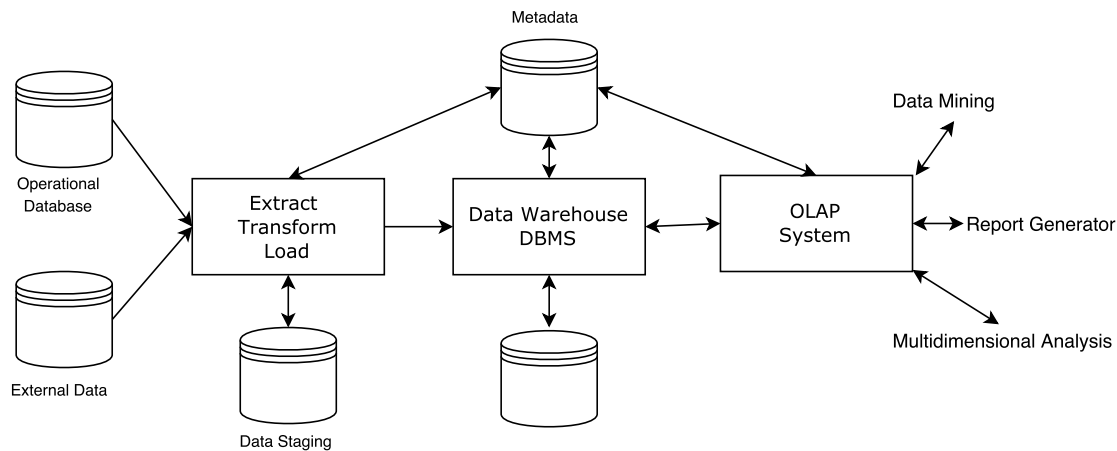
### 3.1 Il processo di *data warehousing*

Con il termine *data warehouse* si intende una particolare base di dati con le seguenti principali caratteristiche [Albano,2014]:

- **Tempificata.** Contiene informazioni sul tempo in cui si verificano gli eventi di interesse. Questa caratteristica permette di realizzare la storicizzazione delle informazioni.
- **Integrata.** I dati non provengono in genere da un'unica sorgente (una base di dati dell'organizzazione), ma sono il risultato di un lungo e costoso processo di integrazione di dati eterogenei.
- **Statica.** I dati vengono usati interattivamente per operazioni di ricerca e non di modifica. Periodicamente ai dati disponibili se ne aggiungono di nuovi o si rimuovono quelli ritenuti obsoleti.
- **Organizzata per soggetti.** Nei sistemi operazionali i dati sono organizzati per eseguire le attività aziendali quotidiane, mentre nei sistemi direzionali i dati sono organizzati per analizzare dei soggetti di interesse che influenzano l'andamento complessivo dell'azienda. Quando i dati riguardano un solo soggetto di interesse si parla di *data mart* e possono essere un sottoinsieme di un *data warehouse* più generale.
- **Finalizzata ad analisi di supporto alle decisioni.** I dati sono organizzati per facilitare le loro analisi al fine di produrre delle opportune sintesi di supporto ai processi decisionali.

Con il termine *data warehousing* si indica il processo attraverso cui i dati sono organizzati in un *data warehouse* e quindi messi a disposizione degli utenti finali utilizzando applicazioni di *business intelligence*. Un'applicazione di questo tipo deve permettere la ricerca di aspetti interessanti deducibili dai dati attraverso strumenti di analisi.

Nel caso qui analizzato, i dati sono organizzati secondo quella che in [Albano,2014] è definita come architettura a tre livelli (Fig. 3.1).



**Figura 3.1:** *Data warehousing* con architettura a tre livelli

Come mostrato in figura 3.1, il processo di *data warehousing* è composto da quattro fasi principali:

- 1) **Sorgenti.** Definizione delle sorgenti dei dati, base di dati operativa ed eventualmente fonti esterne.
- 2) **Extraction transformation loading.** Fase in cui i dati provenienti dalle sorgenti vengono opportunamente estratti, modificati e caricati nel sistema per *data warehouse*.
- 4) **Data warehouse.** Sistema in cui sono opportunamente memorizzati i dati con lo scopo di ottimizzare la loro successiva analisi.
- 5) **Strumenti di analisi.** Insieme di strumenti che interrogando il *data warehouse* permettono di creare reportistica ad-hoc, analizzare i dati in modo multidimensionale ed effettuare operazioni di *data mining*.

Come è noto, questa soluzione garantisce diversi vantaggi:

- separazione tra il DBMS (*database management system*) operativo e quello finale, che potrà essere un sistema progettato appositamente per il supporto alle decisioni;
- separazione tra le applicazioni transazionali e quelle di *business intelligence*, eliminando influenze reciproche sulle prestazioni;
- separazione tra il processo di estrazione e integrazione delle fonti di dati e il processo di riorganizzazione e caricamento dei dati stessi nel *data warehouse* (vero vantaggio aggiuntivo che si ottiene passando dall'architettura a due livelli a quella a tre).

Una volta creato e popolato, il *data warehouse*, necessiterà di fasi di aggiornamento effettuate ad intervalli di tempo regolari per garantire l'allineamento con la base di dati operativa.

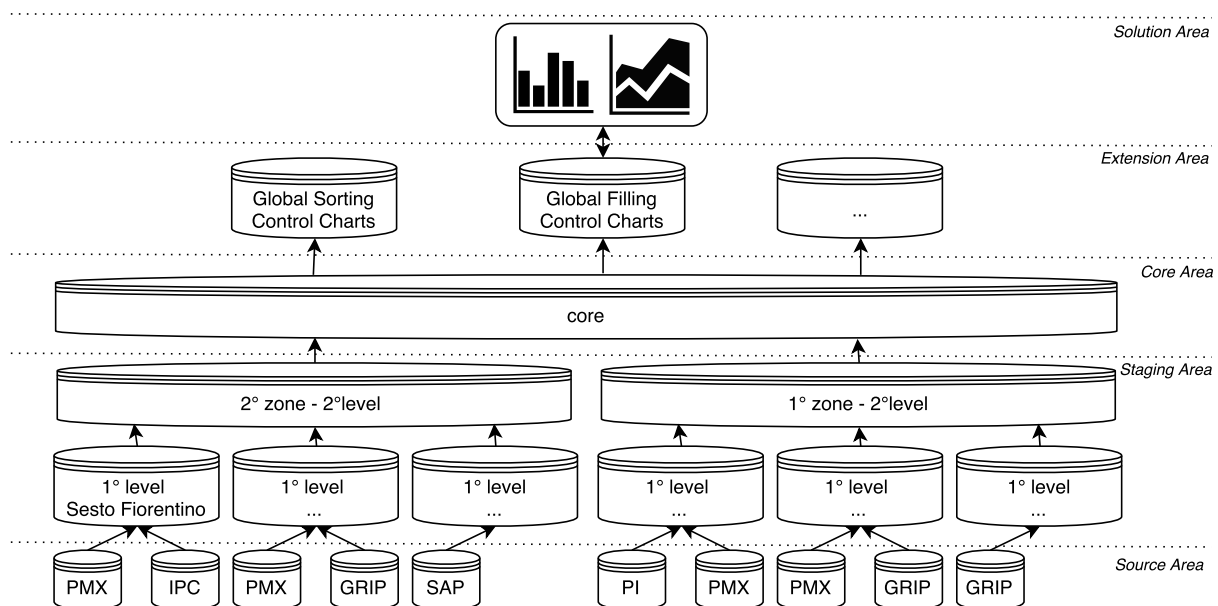
### 3.2 Architettura *global manufacturing data mart*

L'architettura multilivello adottata dalla società cliente prende il nome di "*global manufacturing data mart*" (d'ora in poi GMDM).



Questa architettura (Fig. 3.2), progettata e implementata da SDG Group è nata per soddisfare, oltre le caratteristiche precedentemente descritte, anche le seguenti proprietà:

- **Globale.** Deve permettere l'integrazione di sorgenti differenti provenienti dai diversi siti produttivi.
- **Estendibile.** Deve essere possibile accogliere nuove applicazioni e sistemi sorgenti senza riprogettare integralmente il sistema.
- **Scalabile.** L'architettura *hardware* e *software* deve poter essere facilmente ridimensionata a fronte della crescita nel tempo dei volumi di dati da gestire ed elaborare e del numero di utenti da soddisfare.



**Figura 3.2:** Architettura global manufacturing data mart

Questa architettura, dispone oltre i tre livelli di memorizzazione presentati precedentemente, anche un quarto livello, detto *extension* che si colloca a valle del *data warehouse*, avremo quindi:

- 1) il livello delle sorgenti di dati;
- 2) il livello di *staging*, che contiene i dati ottenuti dal processo di integrazione delle varie sorgenti di dati;
- 3) il livello del *data warehouse* (chiamato *core*);
- 4) il livello dell'estensioni (chiamato *extension*), che contiene i *data mart*.

Questa architettura, integra i dati provenienti da quattro zone (dette anche regioni) differenti:

- zona 1: Stati Uniti, Canada, Mexico e Sud America;
- zona 2: Europa, Medio Oriente e Africa;
- zona 3: Asia Pacifica;
- zona 4: Porto Rico.

### 3.2.1 Source area

Nello strato inferiore del GMDM, troviamo i sistemi sorgenti costituiti da basi di dati operazionali. Le sorgenti sono specifici per sito produttivo.

I sistemi sorgenti più significativi sono quelli riportati nella tabella 3.1.

Area	Sistemi sorgenti
ERP ( <i>enterprise resource planning</i> )	SAP ( <i>systems applications and products</i> )
Quality and compliance	Trackwise
LIMS ( <i>laboratory information management system</i> )	Darwin, VCoM
Maintenance	Maximo
Data historian	PI ( <i>process information</i> )
MES ( <i>manufacturing execution system</i> )	PMX ( <i>production management system</i> ), Syncade
Document management	Regulus
Process automation	GRIP ( <i>global resource integration platform</i> )
In process control	IPC Book ( <i>in process checks</i> )

**Tabella 3.1:** Principali sistemi sorgenti del GMDM

### 3.2.2 Staging area

Data la consistente quantità di dati e tabelle coinvolte nel GMDM, è presente una struttura intermedia tra le sorgenti ed il *core* detta *staging area*.

L'area di *staging* è un'area tecnica per rendere più semplice l'alimentazione del *core* vero e proprio. In questo livello vengono memorizzate le tabelle provenienti dalla fonte senza implementare alcuna trasformazione. Quindi il livello di dettaglio sarà speculare alle informazioni provenienti dai sistemi sorgenti alimentanti.

A differenza della maggior parte di architetture di *data warehousing*, questa area si compone di due livelli. Si ha un primo livello di *staging*, che conterrà solo le tabelle sorgenti locali, e un secondo livello di *staging* che conterrà tutte le tabelle sorgenti appartenenti alla stessa regione.

Ogni giorno, il primo livello di *staging* viene caricato in modalità "*full*", dopo aver effettuato la truncate di tutte le tabelle.

Il secondo livello di *staging* serve per storicizzare l'andamento dei dati nel tempo. Questo livello infatti:

- inserisce i nuovi *record* registrati sul primo livello;
- aggiorna i *record* che hanno subito una variazione ad almeno uno dei campi;
- cancella logicamente eventuali *record* che sono stati eliminati fisicamente sul primo livello.

Entrambi i livelli sono caricati tramite una *store procedure* parametrica scritta in PL/SQL invocata tramite un flusso unico per sorgente e sito produttivo.

Per garantire la globalità del sistema, le tabelle di *staging* contengono i seguenti principali attributi tecnici:

- FCLTY\_CD: è il codice univoco che identifica il sito produttivo del *record*;
- SRC\_INST\_CD: è il codice univoco che identifica l'installazione della sorgente nel sito produttivo;

- RCRD\_STS\_CD: identifica lo stato di un *record*, "A" se il *record* è attivo o "D" se il *record* è stato cancellato;
- INS\_TMSTMP: identifica il *timestamp* di inserimento del *record*;
- UPD\_TMSTMP: identifica il *timestamp* di aggiornamento del *record*;
- END\_TMSTMP: identifica il *timestamp* di cancellazione del *record*.

Degli attributi tecnici sopra elencati le tabelle di *staging* di primo livello contengono solo il "FACILITY\_CD".

A differenza delle tabelle di *staging* di primo livello che non hanno definita la chiave primaria, le tabelle di secondo livello mantengono oltre le chiavi primarie della sorgente, anche i campi tecnici "FACILITY\_CD", "SRC\_INST\_CD", "RCRD\_STS\_CD" e "END\_TMSTMP". Così, è possibile distinguere nello *staging* di secondo livello, *record* della stessa sorgente ma appartenenti a siti produttivi differenti ma della stessa regione.

### 3.2.3 Core area

A partire dall'area di *staging* è possibile alimentare il *core*. Questa è la prima area globale, dove è possibile ospitare le trasformazioni, seguendo le regole che definiscono gli oggetti di *business* necessari alle analisi.

In questa area oltre a mantenere il massimo livello di dettaglio disponibile (i raggruppamenti non sono concessi), si garantisce la normalizzazione su ogni tabella.

Ogni tabella di questa area verrà aggiornata in modalità "*delta*" su riferimento del *timestamp*, mentre una volta a settimana in modalità "*full*".

Non è necessario conservare la storia, quindi i *record* non aggiornati verranno rimossi tramite una *store procedure* scritta in PL/SQL, invocata dopo l'aggiornamento di ogni tabella del *core*.

Nel *core* è possibile definire:

- tabelle di configurazione: usate nelle logiche di caricamento delle tabelle del *core*;
- tabelle di scarto: una per ogni tabella del *core*, che contengono *record* volutamente filtrati dai flussi di caricamento non rilevanti per le analisi o i *record* cancellati dalla procedura di *delete*;
- tabelle dei fatti;
- tabelle delle dimensioni.

Se possibile, ogni tabella nel *core* manterrà la propria chiave di *business* altrimenti vengono definite delle chiavi surrogate.

Pure in questa area, vengono definiti su ogni tabella degli attributi tecnici:

- FACILITY\_CD: è il codice univoco che identifica il sito produttivo del *record*;
- RCRD\_STS\_CD: identifica lo stato di un *record*, "A" se il *record* è attivo o "D" se il *record* è stato cancellato;
- SRC\_CODE: identifica la sorgente del *record*;
- MDM\_LOAD\_ID: è il codice univoco del caricamento;
- MDM\_TRNSCTN\_TMSTMP: identifica il *timestamp* dell'ultimo aggiornamento del *record*.

### 3.2.4 *Extension area*

L'area dell'*extension* è il luogo dove i dati vengono esposti in maniera chiara e compatibile con le regole e gli oggetti di *business*, tale strato, globale, è accessibile agli utenti finali attraverso gli strumenti di *querying* e di *reporting*, ed è costituito da un raggruppamento di "porzioni logiche" dei dati disponibili detti *data mart*.

I *data mart*, sono quindi tabelle denormalizzate usate per memorizzare dati aggregati al fine di fornire informazioni precalcolate e di migliorare le *performance* delle *query* usate nei *tool* di *reporting*.

Ogni *data mart* è aggiornato in modalità "full" e, pure qui i *record* non aggiornati vengono eliminati tramite una procedura di *delete*.

Le tabelle di questo livello conterranno gli stessi attributi tecnici di quelle del *core*.

Peculiarità di questa architettura è che ogni *data mart*, popolato dal *core*, potrà contenere informazioni provenienti dalla stessa (o diversa) sorgente, appartenenti a regioni distinte, ma finalizzati a soddisfare lo stesso bisogno di analisi.

Nell'area dell'*extension* oltre ai *data mart* è possibile definire:

- tabelle di configurazione: usate nelle logiche di caricamento delle tabelle dell'*extension*;
- tabelle delle dimensioni.

Affinché le procedure di *delete* non vadano in errore per "vincolo di integrità" nell'area del *core* e dell'*extension* non vengono definite le chiavi esterne. Questa scelta deriva pure dal fatto, che la definizione di chiavi esterne possono causare un calo di prestazioni significativo durante i caricamenti delle tabelle.

### 3.2.5 *Nomenclatura*

Al fine di garantire un'elevata manutenibilità, il nome di ogni oggetto all'interno del GMDM segue una rigida convenzione che permette di identificarne il tipo e la destinazione d'uso all'interno dell'architettura.

Le principali convenzioni utilizzate sono sintetizzate di seguito:

- *staging area*
  - TBSTMD\_<source\_system>\_<nome\_tabella>: per tabelle dimensionali;
  - TBSTFT\_<source\_system>\_<nome\_tabella>: per tabelle dei fatti.
- *core e extension area*
  - MDM\_<nome\_tabella>\_DM: per tabelle *data mart*;
  - MDM\_<nome\_tabella>\_DIM: per tabelle dimensionali;
  - MDM\_<nome\_tabella>\_CNF: per tabelle di configurazione;
  - MDM\_<nome\_tabella>: per tabelle dei fatti.

Anche gli attributi devono seguire una nomenclatura ben definita. Le regole di seguito, valgono per l'area di *staging*, *core* e *extension*:

- <nome\_campo>\_CD: *varchar*, campo utilizzato come chiave di *business*;
- <nome\_campo>\_ID: *integer*, intero utilizzato per le chiavi surrogate;
- <nome\_campo>\_NBR: *integer*, campo utilizzato per un numerico intero;

- <nome\_campo>\_DT: *date*, campo utilizzato per una data;
- <nome\_campo>\_FLG: *integer/varchar*, campo utilizzato come *flag*;
- <nome\_campo>\_VAL: *decimal*, campo utilizzato per un numerico con virgola;
- <nome\_campo>\_DESC: *varchar*, campo utilizzato per un valore di tipo stringa.



## Capitolo 4

# ANALISI DEI REQUISITI E PROGETTAZIONE CONCETTUALE INIZIALE

Nel presente capitolo dopo aver introdotto i soggetti coinvolti, si presentano le fasi del processo di progettazione del *data mart*. Più specificamente nel capitolo sono descritti, nell'ordine, il modello seguito per lo sviluppo e le fasi di "analisi dei requisiti" e "progettazione concettuale iniziale".

Si sottolinea che le fasi qui descritte si riferiscono esplicitamente a quelle presenti nel modello proposto in [Albano,2014], modello seguito per lo sviluppo della soluzione e per la stesura di questo documento.

Si individuano in tale modello cinque fasi:

- analisi dei requisiti;
- progettazione concettuale iniziale del *data mart* guidata dall'analisi;
- progettazione concettuale candidata guidata dai dati;
- progettazione concettuale finale;
- progettazione logica del *data mart* e del *data warehouse*.

### 4.1 Analisi dei requisiti

La raccolta dei requisiti è stata portata avanti attraverso numerosi incontri preliminari con il cliente, in seguito ai quali è stato realizzato un primo documento con il fine di raccogliere le esigenze di utilizzo del *data mart* e di delineare quali informazioni sono di interesse strategico. Successivamente il documento è stato rivisto e integrato con parti più tecniche, specifiche tecniche più dettagliate, espresse tramite tabelle, esempi di visualizzazione grafica ecc.

I soggetti che hanno preso parte alle riunioni sono stati:

- dipendenti con profili funzionali e/o tecnici di SDG Group;
- dipendenti della direzione centrale sistemi informativi della società cliente;
- futuri utenti del sistema, costituiti da dipendenti della società cliente.

#### 4.1.1 Specifica dei requisiti

Nel corso degli incontri con il committente sono stati individuati alcuni requisiti che la soluzione è tenuta a soddisfare. Al fine di comprendere meglio questa sezione, è utile chiarire il concetto di *batch*.

Un *batch* è un lotto accomunato dallo stesso codice materiale di semifinito, prodotto impiegando una ricetta ed una BOM (*bill of material*) specifica. Nel sito produttivo di Sesto Fiorentino, la linea 1 e 2 produce lotti da 2500 litri, cioè 750000 cartucce. La linea 3 produce lotti da 1250 litri, cioè la metà di cartucce.

Per la soluzione proposta, è importante conoscere:

- il codice univoco del lotto;
- il codice univoco del materiale che identifica il tipo di prodotto che compone un lotto;
- la dimensione del lotto in litri;
- la line su cui è stata effettuata l'attività di *filling*.

Allo scopo di ottenere una visione maggiormente dettagliata per il controllo di qualità, le analisi non riguarderanno solo il singolo *batch*, ma anche i singoli campioni (di dimensioni differenti) che compongono il *batch*.



Di seguito sono riportati i requisiti di analisi più importanti emersi (Tab 4.1).

N.	Requisito di analisi	Dimensioni	Misure	Metriche
1	Media del parametro per tipo di parametro, per materiale del <i>batch</i> , per linea di <i>filling</i> , per dimensione della cartuccia del <i>batch</i> , relativo a lotti riempiti o realizzati in un certo intervallo di tempo.	<i>Parameter (Category, Type), Material (Material code, Description, Size), Phase (Filling line, Filling date)</i>	Valore parametro	AVG (valore parametro)
2	Deviazione <i>standard</i> del parametro per tipo di parametro, per materiale del <i>batch</i> , per linea di <i>filling</i> , per dimensione della cartuccia del <i>batch</i> , relativo a lotti riempiti o realizzati in un certo intervallo di tempo.	<i>Parameter (Category, Type), Material (Material code, Description, Size), Phase (Filling line, Filling date)</i>	Valore parametro	STDDEV (valore parametro)
3	Valore del parametro per <i>batch</i> o per campione del <i>batch</i> relativo a lotti riempiti o realizzati in un certo intervallo di tempo.	<i>Parameter (Category, Type), Phase (Filling date), Batch (Batch code, Sample code, Manufacturing date)</i>	Valore parametro	
4	$C_{pk}$ per <i>batch</i> o per campione del <i>batch</i> relativo a lotti riempiti o realizzati in un certo intervallo di tempo.	<i>Parameter (Category, Type), Phase (Filling date), Batch (Batch code, Sample code, Manufacturing date)</i>	Valore parametro	$C_{pk}$

**Tabella 4.1:** Processo analisi qualità

Come affermato in [Albano,2014] la granularità del singolo fatto determina il tipo di analisi che possono essere effettuate sui dati; in generale è preferibile scegliere una granularità fine in modo da rendere possibili eventuali analisi dettagliate.

Nel nostro caso gli utenti sono interessati alle analisi di qualità effettuate sui singoli campioni o lotti (Tab. 4.2).

Descrizione	Dimensioni preliminari	Misure preliminari
Il fatto descrive la misurazione di uno specifico parametro, effettuato da uno specifico strumento, su un campione di dimensione variabile di un determinato lotto, riempito in un determinato momento.	<i>Batch, Phase, Material, Parameter</i>	Valore parametro

**Tabella 4.2:** Fatto analisi qualità

Quindi, la granularità del fatto è rappresentato dai singoli campioni di dimensione variabile che compongono il batch. Il valore del parametro di un campione, non sarà altro che la media totale del valore del parametro di ciascuna cartuccia appartenente a quel campione. Tuttavia la misura del fatto, non è considerata calcolata, in quanto il livello di dettaglio è già definito nel sistema sorgente.

Di seguito si descrivono le dimensioni specificando per ognuna di esse il nome, una descrizione e la granularità (Tab. 4.3).

Nome	Descrizione	Granularità
<i>Batch</i>	Il <i>batch</i> su cui viene effettuato il controllo di qualità.	Un campione
<i>Phase</i>	Fase di riempimento.	Una fase
<i>Material</i>	Rappresenta il materiale del lotto.	Un materiale
<i>Parameter</i>	La tipologia di parametro a cui si riferisce la misurazione.	Un parametro

**Tabella 4.3:** Dimensioni

Per ogni dimensione si elencano gli attributi e una breve descrizione.

Attributo	Descrizione
<i>Sample code</i>	Identificativo del campione.
<i>Batch code</i>	Identificativo del lotto.
<i>Planned sample</i>	Numero pianificato di cartucce da campionare.
<i>Effective sample</i>	Numero effettivo di cartucce campionate.
<i>Manufacturing date</i>	Data di manifatturazione del lotto (formato DD-MON-YY HH24.MI.SS).

**Tabella 4.4:** *Batch*

Attributo	Descrizione
<i>Filling line</i>	Linea di riempimento.
<i>Filling date</i>	Data di riempimento (formato DD-MON-YY HH24.MI.SS).

Tabella 4.5: *Phase*

Attributo	Descrizione
<i>Material code</i>	Identificativo del materiale.
<i>Description</i>	Descrizione del materiale.
<i>Size</i>	Dimensione del lotto in litri.

Tabella 4.6: *Material*

Attributo	Descrizione
<i>Category</i>	Identificativo della categoria del parametro.
<i>Instrument</i>	Identificativo dello strumento usato per la misurazione.
<i>Unit of measure</i>	Identificativo dell'unità di misura del parametro.
<i>Type</i>	Identificativo della tipologia del risultato.

Tabella 4.7: *Parameter*

Di seguito vengono presentate le gerarchie dimensionali, specificando per ogni dimensione le dipendenze funzionali esistenti tra gli attributi dimensionali e il loro tipo (Tab. 4.8).

Dimensione	Descrizione	Tipo di gerarchia
<i>Batch</i>	<i>Sample code</i> → <i>Batch code</i>	Bilanciata
<i>Parameter</i>	<i>Category</i> → <i>Instrument</i>	Bilanciata

Tabella 4.8: Gerarchie dimensionali

Per quanto riguarda il tipo di strategia utilizzato per trattare le dimensioni con attributi che possono cambiare nel tempo, come riportato in [Albano,2014], le tipologie di trattamento delle modifiche possibili sono quattro:

- **Tipo 1 - perdita di storia.** Il valore dell'attributo dimensionale che cambia viene sostituito con il nuovo valore; è la soluzione più semplice ed immediata, ma si perde la possibilità di storicizzare i cambiamenti.
- **Tipo 2 - conservazione della storia.** Si aggiunge una nuova riga alla tabella dimensionale, creando di fatto una entità nuova. Tutti i fatti precedenti alla modifica fanno riferimento alla vecchia entità, mentre tutti i fatti successivi alla modifica fanno riferimento a quella nuova. In questo modo il caricamento dei dati si complica, aumentano i dati della dimensione, ma si ha la possibilità di storicizzare il cambiamento.
- **Tipo 3 - conservazione di una o più versioni della storia.** Con questo trattamento si memorizza, oltre che alla storia, anche il momento temporale in cui avviene il cambiamento.

- **Tipo 4 - modifiche con alta frequenza.** Per gli attributi dimensionali che cambiano molto frequentemente si possono prevedere due tabelle dimensionali, una contenente gli attributi che rimangono immutati e una contenente gli attributi che variano.

Nel nostro caso, come spiegato precedentemente (Sez. 3.2), nel livello del *core* e dell' *extension*, non interessa la storizzazione dei cambiamenti e quindi come riportato in tabella 4.9, tutte le dimensioni sono trattate con il "tipo 1".

Nome	Granularità	Trattamento
<i>Batch</i>	Un campione	Tipo 1
<i>Phase</i>	Una fase	Tipo 1
<i>Material</i>	Un materiale	Tipo 1
<i>Parameter</i>	Un parametro	Tipo 1

**Tabella 4.9:** Trattamento delle modifiche

Nella tabella 4.10 viene descritta l'unica misura rilevante per le analisi, la sua aggregabilità e la formula nel caso di misura calcolata.

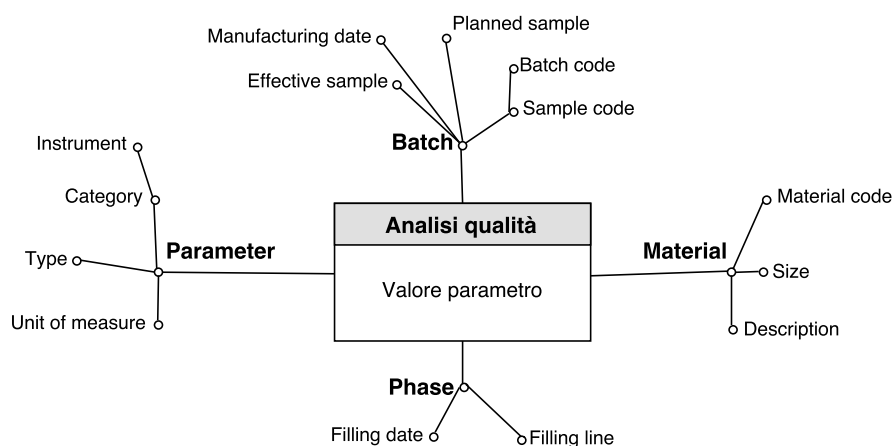
Misura	Descrizione	Aggregabilità	Calcolata
Valore parametro	Valore del parametro di un determinato campione di un <i>batch</i> .	Additiva	No

**Tabella 4.10:** Misura

## 4.2 Progettazione concettuale iniziale

La raccolta dei requisiti e la traduzione formale di questi ha permesso di identificare requisiti specifici sul fatto, sulle misure associate e sulle dimensioni.

È possibile esprimere e rappresentare graficamente una sintesi delle informazioni raccolte attraverso il formalismo DFM (*dimensional fact model*). Il modello ottenuto come risultato dall'analisi dei requisiti è presentato nella figura 4.1.



**Figura 4.1:** Modello concettuale iniziale

## Capitolo 5

# PROGETTAZIONE CONCETTUALE FINALE E LOGICA

Nel presente capitolo, sono affrontate le ultime fasi relative al *design* del *data mart* secondo il modello adottato. La prima di queste fasi riguarda la progettazione concettuale candidata, cioè quella ottenuta a partire dall'osservazione delle fonti di dato disponibili, che richiede pertanto una presentazione iniziale del sistema operativo di partenza e delle fonti già presenti nel *core*. La successiva fase, quella di progettazione concettuale finale, prevede un affinamento ed un'integrazione dei due modelli preliminarmente ottenuti, allo scopo di comprendere e colmare le eventuali differenze riscontrate in essi. Infine, una volta ottenuto il modello concettuale finale, si procede alla presentazione della progettazione logica che, in questo contesto specifico, si sostanzia nella traduzione in un modello relazionale degli aspetti precedentemente modellati.

Al fine di comprendere meglio, il modello logico finale descritto nell'ultima parte di questo capitolo, è indispensabile definire, oltre ai requisiti di analisi descritti nella sezione 4.1; anche i seguenti requisiti funzionali che la soluzione *global filling control charts* dovrà soddisfare:

- I valori dei limiti delle carte di controllo, non saranno estratti dal sistema sorgente, ma saranno definiti in una tabella di configurazione. Tramite questa tabella è possibile settare il valore dei limiti e la visibilità dei limiti per ogni singolo parametro.
- I materiali lavorati in ogni linea di riempimento, saranno definiti in una tabella di configurazione.
- I materiali che fanno parte dell'intero processo produttivo, saranno definiti in una tabella di configurazione.

L'uso delle tabelle di configurazione è utile al fine di agevolare, eventuali richieste di cambiamento da parte del cliente. Se un particolare materiale, è richiesto per una particolare soluzione piuttosto che per un'altra, basterà settare il rispettivo *flag* di visibilità nelle tabelle di configurazione, senza andar a intaccare la logica dei flussi di integrazione.

### 5.1 Presentazione delle fonti di dato disponibili

Le sorgenti che alimentano il sistema GMDM sono molteplici e per poter realizzare quanto descritto nella modellazione concettuale è stato necessario integrare e trasformare i dati con un processo molto dispendioso, come descritto meglio nel Capitolo 7.

La fonte principale di dati per la realizzazione del *data mart* su cui poggia la soluzione *global filling control charts* è l' "IPC Book". E' stato lavoro di questo progetto, integrare la sorgente IPC Book nel GMDM, al fine di rendere disponibili i dati per la nuova soluzione.

### 5.1.1 Fonte IPC Book

Il sistema informativo IPC Book (*in process checks*) fornito dalla società OSI Software [OSI Software,2016] è usato per collezionare e salvare su *database* Oracle, molteplici informazioni, che rappresentano fattori di controllo del processo di fabbricazione, registrati durante l'intero ciclo di vita della cartuccia di insulina.

Questi dati, sono registrati dai seguenti strumenti:

- DIMCARP (*dimensional check on cartridges*): per il controllo delle dimensioni delle cartucce;
- UPG (*universal punch gauge*): mette alla prova diverse dimensioni delle cartucce:
  - 1) *cartridge glass body diameter*: il diametro del corpo di vetro della cartuccia;
  - 2) *combined disc seal diameter*: è valutata misurando il diametro del cilindro più piccolo (coassiale con il corpo) che contiene il punto sulla guarnizione che è più lontano dall'asse del corpo;
  - 3) *disc seal coaxiality*: è valutata misurando il diametro del cilindro più piccolo (coassiale con il corpo) che contiene sia il centro del profilo di tenuta della guarnizione e il centro del corpo cartuccia;
  - 4) *enveloped seal disc diameter*: è valutata misurando il diametro del cilindro più piccolo (in direzione, ma non coassiale con il corpo) che avvolge il profilo della guarnizione.
- ZWICK: per le prove di trazione, compressione e flessione.

I risultati delle misurazioni possono indicare se è necessaria un'azione correttiva per mantenere il processo o il prodotto entro gli intervalli stabiliti, al fine di verificare la conformità dimensionale e meccanica complessiva delle cartucce agli *standard* qualitativi della società cliente.

Per la realizzazione di questa soluzione, è stata resa disponibile una vista (con nome TBSTFT\_IPC\_STATS\_VIEW), contenente i controlli di qualità svolti nella fase di riempimento della cartuccia di insulina. Le informazioni riguardano le misure effettuate su un *batch* di insulina, e sui suoi rispettivi campioni.

Nella tabella 5.1 si descrivono gli attributi principali che compongono la vista.

Attributo	Descrizione
BATCH	Identificativo del lotto.
INSTRUMENT	Identificativo dello strumento usato per la misurazione.
SAMPLING_ID	Identificativo del campione.
MEASURE	Identificativo della categoria del parametro.
CALC	Identificativo della tipologia del risultato.
VALUE	Valore del parametro.
VALUE_DT	Data di elaborazione.
LIM_LL	Limite minimo del risultato.
LIM_UL	Limite massimo del risultato.
NOMINAL	Valore target del risultato.
PRODUCT_CODE	Identificativo del materiale.
CART_PER_PLAN	Numero pianificato di cartucce da campionare.
CART_PRODUCE	Numero effettivo di cartucce campionate.
SAMPLING_TYPE	Tipo di campionamento.

**Tabella 5.1:** Fonte IPC *Book*

La vista, sarà integrata nel *core* come una tabella dei fatti con il nome "MDM\_INSTRUMENT\_MEASURES".

### 5.1.2 Fonte PMX

Dai dati disponibili da IPC *Book* (vedi Tab. 5.1), è possibile notare, che mancano informazioni riguardante la data di riempimento e di manifatturazione del *batch*. Queste informazioni vengono reperite da PMX, già integrata su GMDM.

Il sistema informativo PMX (*production management system*) rilasciato dalla società Rockwell Automation [Rockwell Automation,2016], viene classificato a livello funzionale come MES (*manufacturing execution system*), ovvero un sistema informatizzato che ha la principale funzione di gestire e controllare la funzione produttiva di un'azienda. La gestione coinvolge il dispaccio degli ordini, gli avanzamenti in quantità e tempo, il versamento a magazzino, nonché il collegamento diretto ai macchinari per dedurre informazioni utili ad integrare l'esecuzione della produzione, e delle informazioni per il controllo della produzione stessa.

Le tabelle, utili per la soluzione sono:

- MDM\_PROD\_ORDER: per il reperimento della data di riempimento del *batch*;
- MDM\_BATCH\_DIM: anagrafica del batch, per il reperimento della data di manifatturazione del *batch*.

## 5.2 Progettazione concettuale candidata e finale

L'analisi delle fonti disponibili non stravolge il modello concettuale iniziale, anzi lo conferma per quanto riguarda la maggioranza dei suoi aspetti. Pertanto, il modello concettuale mostrato nella sezione 4.2, viene considerato come il modello concettuale finale.

### 5.3 Progettazione logica del *data mart*

Terminata la fase di progettazione concettuale, si provvede ad individuare lo schema logico risultante.

Questo è reso possibile dopo l'integrazione nel *core* delle seguenti tabelle di configurazione:

- MDM\_PARAMETERS\_LIMITS\_CNF: utilizzata per fornire alcune informazioni (i limiti dei parametri, l'unità di misura dei parametri, i tipi di parametri...) per ogni materiale utilizzato nel processo di produzione;
- MDM\_MATERIAL\_LINES\_CNF: utilizzata per fornire alcune informazioni (il numero della linea di *formulation*, di *filling*, di *sorting*...) per ogni materiale utilizzato nel processo di produzione.
- MDM\_LINE\_OPER\_PHASE\_CNF: utilizzata per fornire la lista degli identificativi dell'operazione per linea di processo.

Il modello logico ottenuto è mostrato nella figura 5.1.

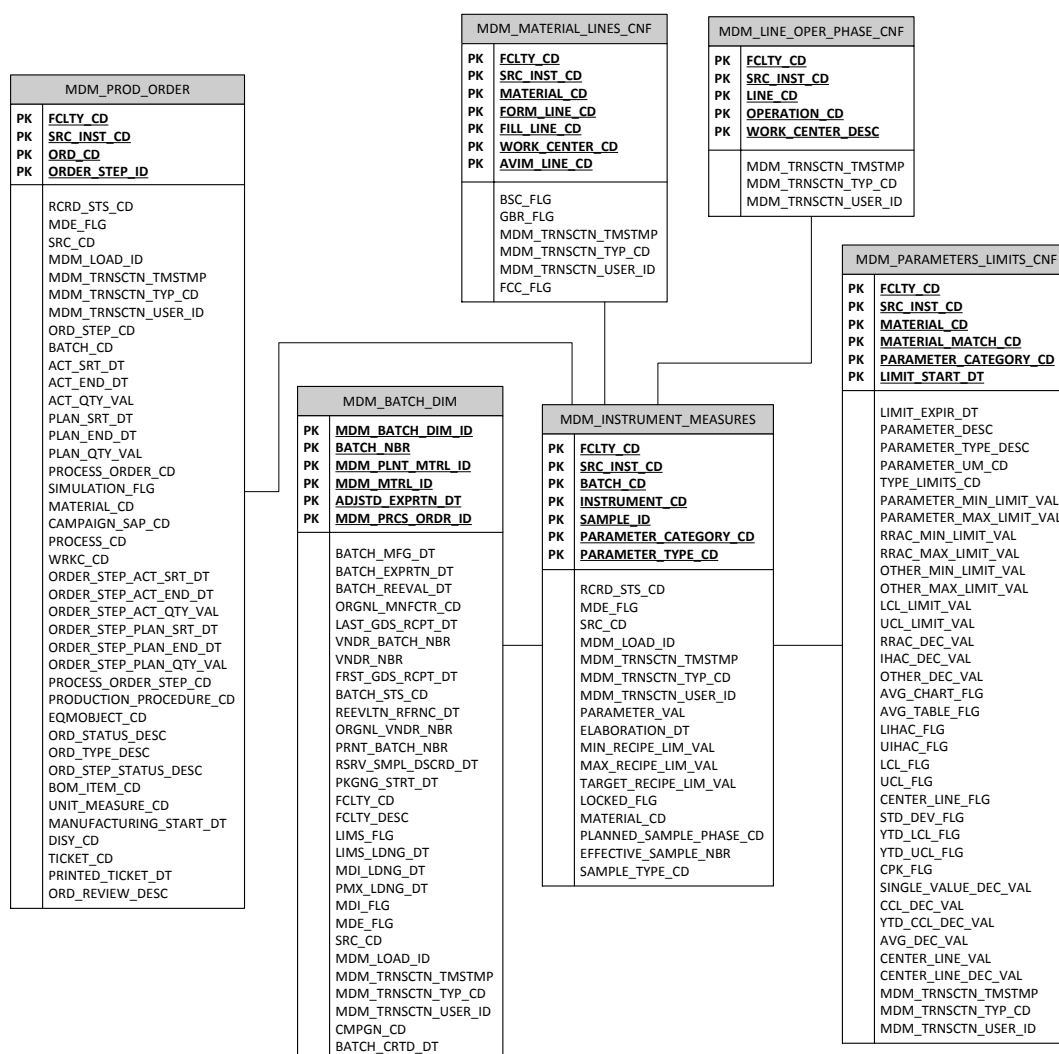


Figura 5.1: Schema logico del *data mart* analisi qualità



Dal modello logico (Fig.5.1), è possibile alimentare l'*extension area* che conterrà la tabella *data mart* (con nome MDM\_CONTROL\_CHART\_FILLING\_DM) della soluzione *global filling control charts* (Tab. 5.2).

Attributo	Descrizione
BATCH_CD	Identificativo del lotto.
FCLTY_CD	Identificativo del sito produttivo.
SRC_CD	Identificativo della sorgente.
MATERIAL_CD	Identificativo del materiale.
OPERATION_CD	Identificativo del processo.
PHASE_CD	Numero pianificato di cartucce da campionare.
OPERATION_INDEX_VAL	Identificativo del campione.
OPERATION_TYPE_DESC	Descrizione del processo.
PARAMETER_CD	Identificativo del parametro.
QUALITY_TYPE_CD	Identificativo della qualità.
OPERATION_DESC	Tipo di campionamento.
PARAMETER_TYPE_CD	Identificativo della tipologia del risultato.
PARAMETER_CATEGORY_CD	Identificativo della categoria del parametro.
PARAMETER_DESC	Descrizione del parametro.
FILLING_LINE_CD	Linea di riempimento.
FILLING_DT	Data di riempimento.
PARAMETER_QUANTITY_VAL	Valore del parametro.
STAT_SAMPLE_SIZE_NBR	Numero effettivo di cartucce campionate.
MANUFACTURING_DT	Data di manifatturazione.

**Tabella 5.2:** Tabella *data mart* della soluzione *global filling control charts*



## Capitolo 6

# AMBIENTE DI SVILUPPO

In questo capitolo verranno descritti i principali strumenti *software* scelti dalla società cliente, utilizzati per le fasi di sviluppo e caricamento del *data warehouse* e realizzazione dell'interfaccia utente.

Si introdurrà il *database management system* (DBMS) Oracle, utilizzato per tutti i database trattati, facendo riferimento a "Oracle Sql Developer" per la gestione e lo sviluppo di soluzioni basate sull'utilizzo del *database* Oracle.

Ci si soffermerà su "Informatica PowerCenter", strumento usato per le operazioni di estrazione, trasformazione e caricamento, mentre per la realizzazione dell'interfaccia utente verrà introdotto "Microsoft SQL Server Reporting Services Report Builder".

A causa della vastità degli argomenti e della non centralità di questi per il lavoro presentato, non si fornirà una descrizione dettagliata di altri strumenti utilizzati quali:

- il linguaggio PL/SQL, estensione di tipo procedurale del linguaggio SQL di proprietà di Oracle, che mette a disposizione, oltre a tutti i costrutti dell'SQL, anche i controlli condizionali, i cicli iterativi, la gestione delle variabili e delle costanti, la gestione delle eccezioni, delle procedure, delle funzioni, ecc;
- la piattaforma "Microsoft Share Point", che consente di aumentare l'efficienza di organizzazioni, team e unità aziendali migliorando la collaborazione tra gli utenti e la condivisione e la gestione di informazioni.

Alcune caratteristiche di questi saranno riprese e spiegate, qualora ritenute importanti per la comprensione dell'elaborato.

### 6.1 Aspetti generali

Ogni tipo di prodotto *software* è controllato da diverse fasi in quello che viene chiamato in letteratura "ciclo di vita del software". All'interno dello stesso si trovano attività che possiamo definire preliminari, in quanto precedono lo sviluppo di codice vero e proprio (analisi e specifica dei requisiti, coinvolgimento degli utenti, pianificazione delle attività, progettazione dell'architettura...), altre incentrate interamente sul codice (implementazione, collaudo e rilascio), oltre a fasi complementari di importanza non trascurabile come documentazione e manutenzione dei sistemi.

La soluzione tradizionalmente applicata per il progetto descritto, prevede differenti ambienti in cui portare avanti le diverse fasi di gestione del software, che sono:

- 1) **Sviluppo**, ambiente in cui si svolge la fase di implementazione, concreta realizzazione della soluzione.

- 2) **Collaudo**, ambiente in cui si verifica e si valida di quanto (misurabilità) il prodotto software implementato soddisfa i requisiti individuati dall'analisi. Il collaudo, in altre parole, valuta la correttezza rispetto alle specifiche.
- 3) **Produzione**, ambiente finale in cui il software, una volta che ha superato con successo le verifiche della fase di collaudo, viene rilasciato al cliente, è eseguito nel contesto reale.

L'ambiente di produzione conclude lo sviluppo e il relativo collaudo e dà inizio alla manutenzione.

Infatti, può capitare che dopo il rilascio in produzione siano segnalati da parte degli utenti dei difetti nel *software* (*bug*) che necessitano correzioni più o meno urgenti, o richieste di modifiche dipendenti da fattori esterni determinando la necessità di un adeguamento. In tal caso le modifiche necessarie sono implementate in sviluppo, testate nuovamente in collaudo, rilasciate in produzione e così via.

Questa struttura è replicata su tutte le parti coinvolte nel progetto (*database*, flussi ETL, reportistica ecc.).

## 6.2 Database Oracle

Oracle Corporation è una notissima multinazionale del software che deve buona parte del suo successo ad un'intuizione della fine degli anni settanta: nel 1979, infatti, l'azienda fondata da Larry Ellison diffuse in commercio il primo *database* relazionale al mondo.

Con i suoi prodotti commerciali, Oracle è tra i primi fornitori di sistemi per *data warehouse* e basi di dati in termini di fatturato, e si posiziona tra i *leader* del mercato nel *magic quadrant* di Gartner [Gartner,2015].

Il sistema utilizzato in questa implementazione è "Oracle Database 11g Release 2", un *database management system* (DBMS) relazionale a oggetti, che supporta le principali funzioni SQL analitiche.

Tra le varie potenzialità è la programmabilità, ossia la possibilità di poter memorizzare ed eseguire procedure e funzioni in linguaggio Oracle PL/SQL.

## 6.3 Oracle SQL Developer

"Oracle SQL Developer" è l'ambiente di sviluppo integrato (IDE) utilizzato per la realizzazione e la gestione del sistema di *data warehouse* su Oracle.

Lo strumento consente a sviluppatori, progettisti e DBA di interfacciarsi con il *database* al fine di interagire con gli oggetti presenti.

Nell'ambito dell'implementazione descritta in questo lavoro si è fatto ricorso alle funzionalità di Oracle SQL Developer per:

- la creazione dell'interfaccia con la base di dati operazionale;
- la creazione e il *debug* delle procedure e delle funzioni in linguaggio PL/SQL;
- l'esecuzione di interrogazioni SQL e la verifica dei piani di accesso;
- la verifica delle prestazioni complessive del sistema;
- la gestione del modello logico e del modello fisico della base di dati.

## 6.4 Informatica PowerCenter

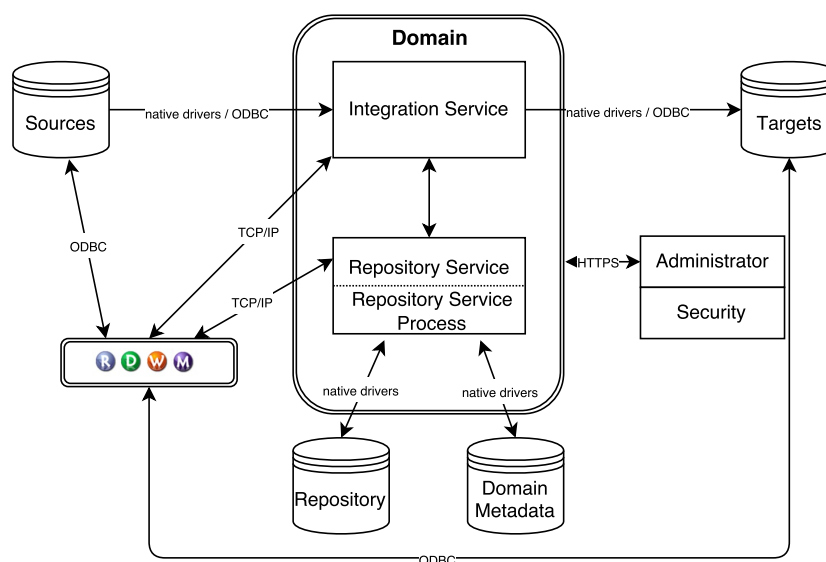
Informatica PowerCenter è uno strumento di ETL per lo sviluppo di *data warehouse* prodotto dalla società Informatica Corporation, considerato il principale fornitore indipendente di *software* per l'integrazione dei dati. Quanto detto è dimostrato dal *magic quadrant* di Gartner che posiziona Informatica come *leader* da 10 anni consecutivi [Gartner,2015].

PowerCenter permette di accedere e quindi estrarre i dati provenienti da diverse fonti eterogenee di aziende e organizzazioni di qualsiasi dimensione, applicando una varietà di trasformazioni secondo la logica di *business* desiderato, per poi integrarli nelle strutture informative predisposte come un *data warehouse*, tutto questo con elevate prestazioni in termini di precisione e tempestività.

Nel realizzare tale sezione si è fatto riferimento a manuali tecnici in particolare [Informatica,2013].

### 6.4.1 Architettura

L'architettura può essere divisa in due macro aree, il dominio e gli strumenti *client* (Fig. 6.1).



**Figura 6.1:** Architettura di Informatica PowerCenter

Il "dominio" è l'unità primaria per la gestione e l'amministrazione all'interno di PowerCenter. È la raccolta di tutti i server necessari per supportare le funzionalità di PowerCenter. Ogni volta che si desidera utilizzare i servizi di PowerCenter si invia una richiesta al server di dominio. *Repository service* e *integration service* sono componenti del dominio di PowerCenter.

Il "*repository service*" gestisce le connessioni tra il client e il repository. È un separato processo *multi-threaded*, che recupera, inserisce, e aggiorna i metadati nelle tabelle del repository. Esso assicura la coerenza dei metadati nel repository.

Quando si avvia un flusso di lavoro (ovvero un insieme di istruzioni che descrive come e quando eseguire compiti legati all'estrazione, alla trasformazione e al caricamento dei dati) parte l'"*integration service*" che legge le informazioni dal repository per recuperare i metadati. Assegna e materializza i valori delle variabili/parametri estratti dalle sorgenti, applica le regole di trasformazioni configurati nel flusso di lavoro (*work flow*) e memorizza i dati. Infine, carica i dati trasformati nelle destinazioni. Questo servizio riesce a combinare dati provenienti da diverse piattaforme e sorgenti eterogenee. Ad esempio, è possibile unire i dati da un *file flat* e una sorgente di Oracle. Inoltre può caricare i dati in diverse piattaforme e destinazioni eterogenee.

Gli strumenti *client* sono utilizzati per gestire gli utenti, definire le origini e le destinazioni e creare flussi di lavoro con logiche di trasformazione.

Fanno parte degli strumenti *client*:

- il *repository manager*, che si occupa delle attività di gestione dei *repository*, assegnando privilegi e autorizzazioni a utenti o gruppi, permettendo di visualizzare i metadati dei *repository*;
- il *designer*, per creare le mappature o flussi dati che contengono le istruzioni di trasformazione per l' *integration service*;
- il *workflow manager*, per creare, gestire, pianificare e eseguire i *work flow*;
- il *workflow monitor*, utilizzato per monitorare i *work flow* schedulati e in esecuzione per ogni *integration service*.

### 6.4.2 Repository manager

Attraverso il *repository manager* è possibile osservare gli oggetti contenuti in un *repository* e svolgere su di essi alcune operazioni fondamentali: ricerca, confronto, analisi delle dipendenze, convalida, creazioni di chiavi di accesso.

Ogni *repository*, navigabile attraverso il *repository manager*, appartiene ad un dominio. Prima di poter analizzare il *repository* è quindi necessario fornire le informazioni per la connessione al dominio.

Una volta effettuato l'accesso a un *repository* è possibile navigare tra gli oggetti dello stesso tipo creati con il *designer* (*source*, *target*, trasformazioni, *mapplet* e *mapping*) o con il *workflow manager* (*task*, *session*, *worklet* e *workflow*) organizzati e archiviati all'interno di cartelle distinte a seconda della tipologia (Fig. 6.2).

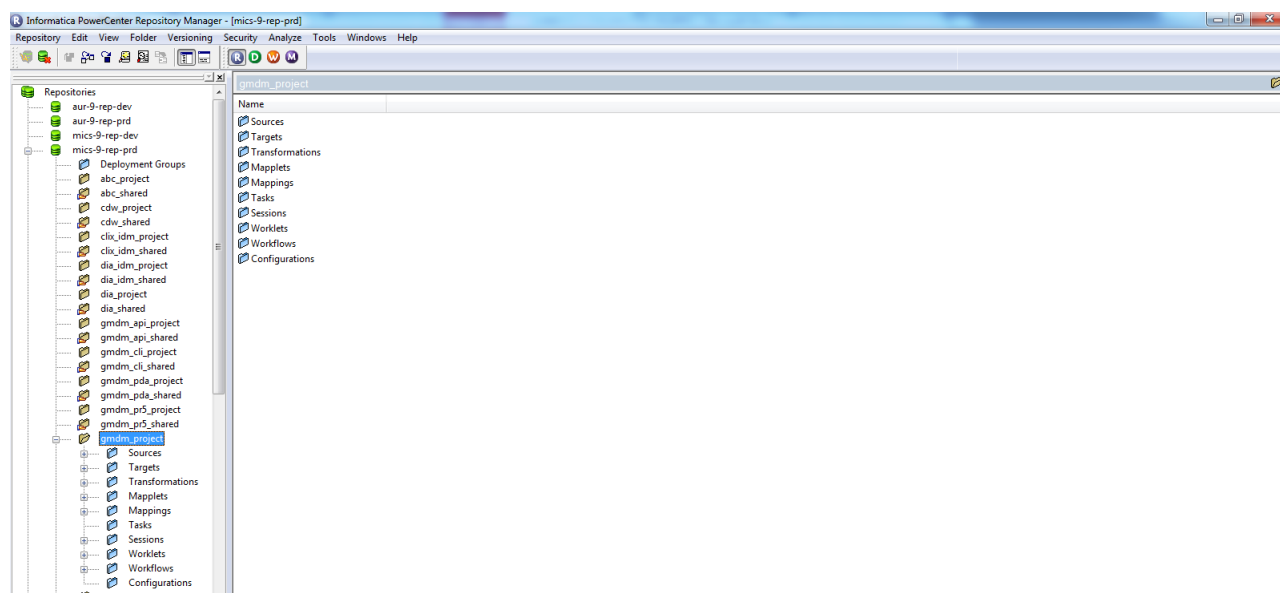
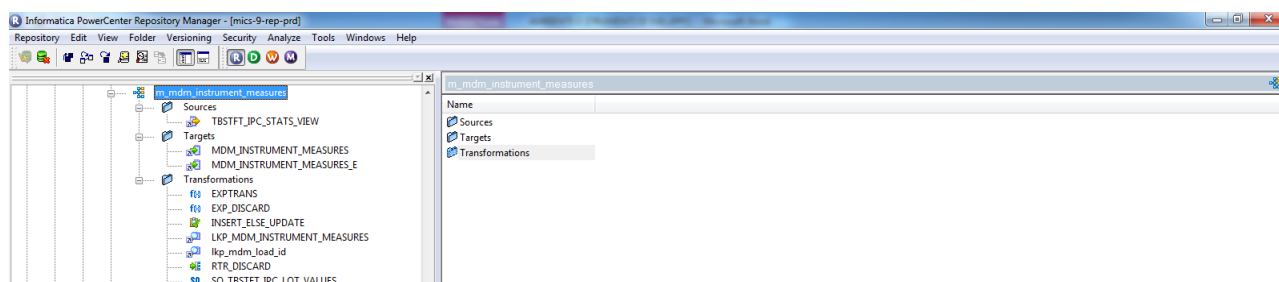


Figura 6.2: Finestra principale del *repository manager*

La finestra principale del *repository manager*, infine, visualizza informazioni riguardanti l'oggetto selezionato nella finestra di navigazione permettendo un'analisi dettagliata delle sue proprietà (Fig. 6.3).



**Figura 6.3:** Finestra principale del *repository manager* con le proprietà dell'oggetto selezionato

Un compito fondamentale del *repository manager*, utilizzato in questo lavoro, è quello della migrazione degli oggetti, per il passaggio nei diversi ambienti di gestione del *software*.

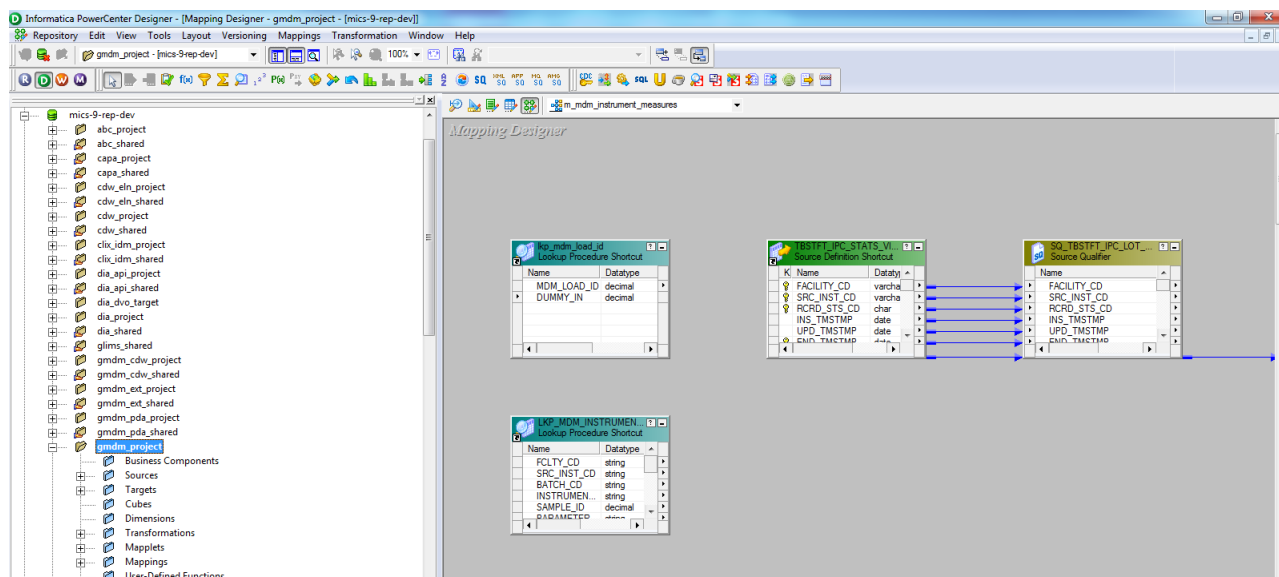
### 6.4.3 Designer

Il *designer* permette di progettare i *mapping* e i *mapplet*.

I *mapplet* è un oggetto riutilizzabile, che contiene una serie di trasformazioni e consente di riutilizzare la logica di trasformazione in più *mapping*.

Attraverso il *designer* è possibile definire la struttura degli oggetti che rispecchiano le tabelle sorgenti, le tabelle cui i dati sono destinati e le trasformazioni a cui i dati devono essere sottoposti prima della loro scrittura nei *database* di destinazione.

Nel *designer* (come per il *repository manager*) è presente la finestra di navigazione laterale, dove è possibile visualizzare gli oggetti che fanno parte del *repository* a cui si è connessi, permettendo di lavorare su più cartelle e *repository* in una sola volta (Fig. 6.4).



**Figura 6.4:** Finestra principale del *designer*

Il *designer* comprende quattro diversi strumenti:

- il *source analyzer*, usato per importare o creare gli oggetti di origine;
- il *target designer*, usato per importare o creare gli oggetti di destinazione;

- il *transformation developer*, usato per la creazione delle trasformazioni riutilizzabili;
- il *mapplet designer*, usato per creare i *mapplet*;
- il *mapping designer*, usato per creare i *mapping*.

Un *mapping* è quindi un insieme di definizioni di origine e di destinazione collegati da oggetti di trasformazione che definiscono le regole per la trasformazione dei dati (Fig. 6.5).

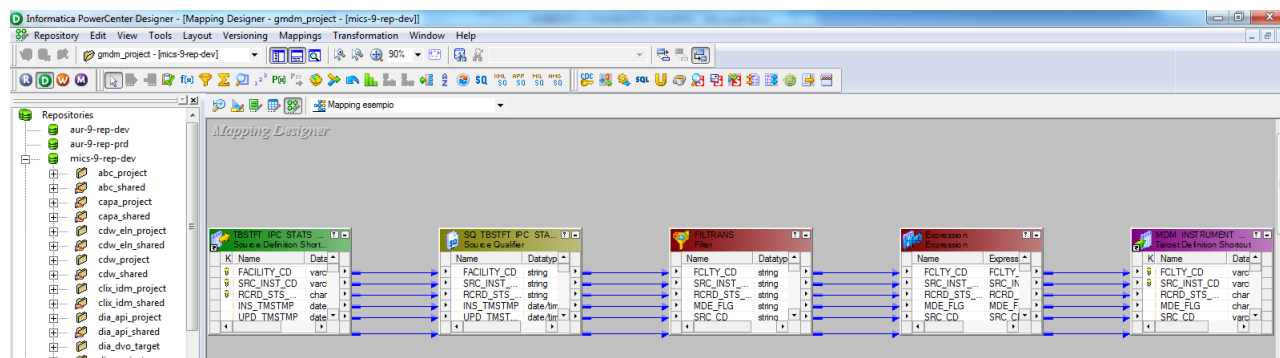


Figura 6.5: Esempio di *mapping* creato con il *designer*

Un *mapping* rappresenta il flusso di dati tra origine e destinazione. Quando l'*integration service* viene eseguito una *session* (definita e creata nel *workflow manager*), utilizza le istruzioni configurate nel *mapping* per leggere, trasformare e scrivere i dati.

Ogni *mapping* deve contenere i seguenti componenti:

- La definizione della sorgente, che descrive le caratteristiche di una tabella di origine o di un *file*.
- Le trasformazioni, per modificare i dati prima della scrittura nella destinazione. Si possono utilizzare diversi oggetti di trasformazione per svolgere funzioni diverse.
- La definizione della *target*, che definisce la tabella o il *file* di destinazione.
- I links, ovvero i collegamenti tra le sorgenti e il *target*, e le trasformazioni in modo che l'*integration service* è in grado di spostare i dati che trasforma. Inoltre, un *mapping* può anche contenere una o più *mapplet*.

È possibile trascinare all'interno della finestra di lavoro tanti oggetti *source* quante sono le sorgenti, tanti oggetti *target* quante sono le destinazioni e tanti altri oggetti quante sono le trasformazioni in cui i dati devono essere sottoposti. Sorgente e destinazione devono essere prima definite rispettivamente nel *source analyzer* e *target designer*.



Qui è possibile importare la struttura di una sorgente o di una destinazione dal *database* che li contiene. Una volta importato il nuovo oggetto, sarà contenuto nel *repository* su cui eravamo collegati all'interno della *folder "target"* (Fig. 6.6).

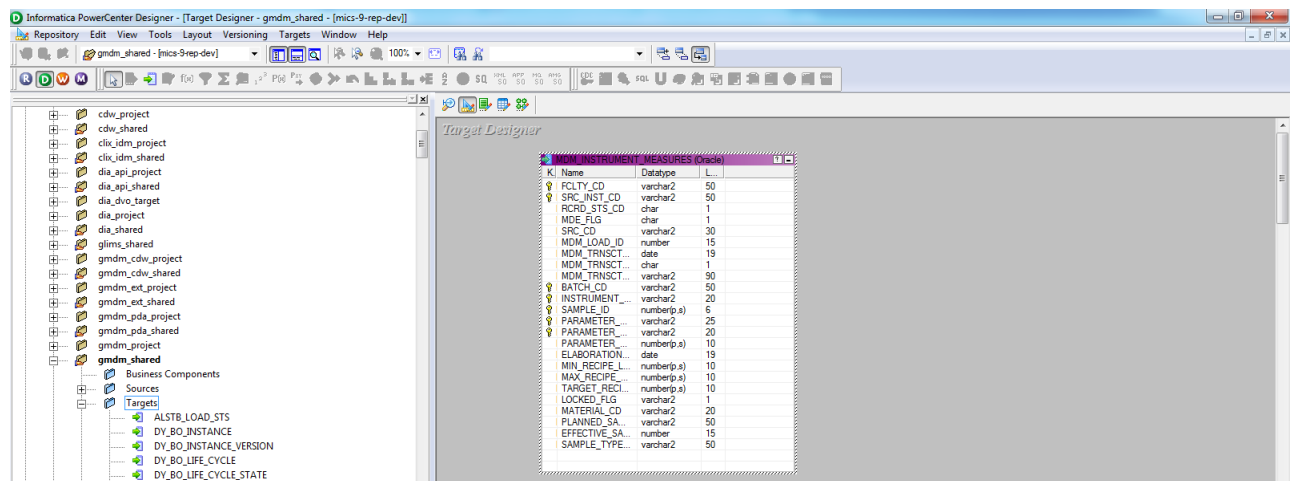


Figura 6.6: Finestra principale del *target designer*

Una volta definiti gli oggetti *source* e *target*, possono essere trascinati nella finestra di lavoro.

Per le sorgenti che provengono da un *database* relazionale è stato abilitato automaticamente il "*source qualifier*" un'opzione con cui è possibile personalizzare la selezione di dati utilizzando una *query SQL* (Fig. 6.5).

Tutte le trasformazioni (le operazioni da effettuare sui dati) vengono progettate con l'ausilio di un'interfaccia grafica che permette di collegare *source*, trasformazioni e *target* mediante l'uso dei "*link*" che collegano le porte corrispondenti ai campi di *input (source)* a quelli di *output (target)*.

Le trasformazioni si dividono in "*passive*", ovvero che non cambiano il numero di righe ricevute e "*attive*", che invece possono cambiare il numero di righe ricevute.

L'elemento caratterizzante delle trasformazioni sono le espressioni. Un'espressione è un'istruzione condizionale o un calcolo che può essere aggiunto a una trasformazione, che consente di modificare, aggiungere o sopprimere singole porte di una singola riga. Aiuta ad attuare le complesse trasformazioni di dati, per applicare logiche di *business* eseguendo controlli e validazioni. I dati possono essere modificati utilizzando gli operatori logici e numerici o le funzioni *built-in*.

Esempio di trasformazioni gestite attraverso le espressioni, possono essere:

- di manipolazione dei dati, come concatenazione, troncamento, arrotondamento (CONCAT, LTRIM, UPPER, INITCAP);
- di conversione (TO\_DECIMAL, TO\_CHAR, TO\_DATE);
- di pulizia dei dati come controllo dei valori *null*, sostituzione dei caratteri, controllo di spazi o per numeri (ISNULL, ReplaceStr);
- di manipolazione di dati di tipo *date*, come conversioni, *test* o aggiunta (GET\_DATE\_PART, IS\_DATE, DIFF\_DATES);
- calcoli scientifici o operazioni numeriche, come esponenziali, logaritmi, moduli, elevamento a potenza (LOG, POWER, SQRT);
- specifici dell'ETL, come operatore *if*, *lookup*, *decode* (IIF, DECODE).

Un'espressione è costituita da porte (*input*, *input/output*, variabili), funzioni, operatori, variabili, valori di ritorno e costanti. Le espressioni possono essere utilizzate nei seguenti oggetti di trasformazione:

- *expression*, trasformazione di tipo passivo per la manipolazione dei dati;
- *aggregator*, trasformazione di tipo attivo per eseguire calcoli di aggregazione;
- *rank*, trasformazione di tipo attivo utilizzato per selezionare il rango superiore o inferiore dei dati;
- *filter*, trasformazione di tipo attivo per eseguire filtri sui dati;
- *update strategy*, trasformazione di tipo attivo per determinare se inserire, cancellare, aggiornare o rifiutare le righe nella tabella target.

Qualora si voglia, ad esempio, concatenare il contenuto di due campi aventi un formato stringa è possibile utilizzare l'oggetto *expression* inserendo l'espressione riportata nella figura 6.7.

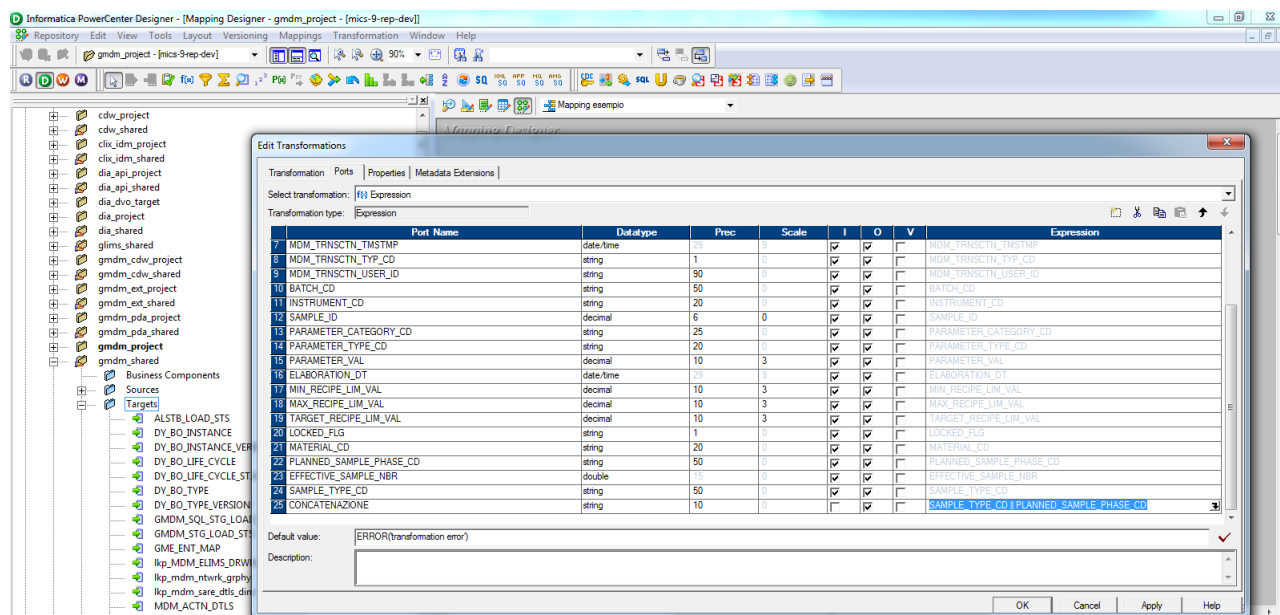


Figura 6.7: Concatenazione di due campi di tipo *string*

Aprendo l'editor dell'oggetto *expression* è possibile visualizzare le funzioni, le porte e le variabili che possono essere usate nell'espressione. Nella finestra a destra è possibile modificare l'espressione e successivamente validarla premendo il tasto "validate". (Fig. 6.8).

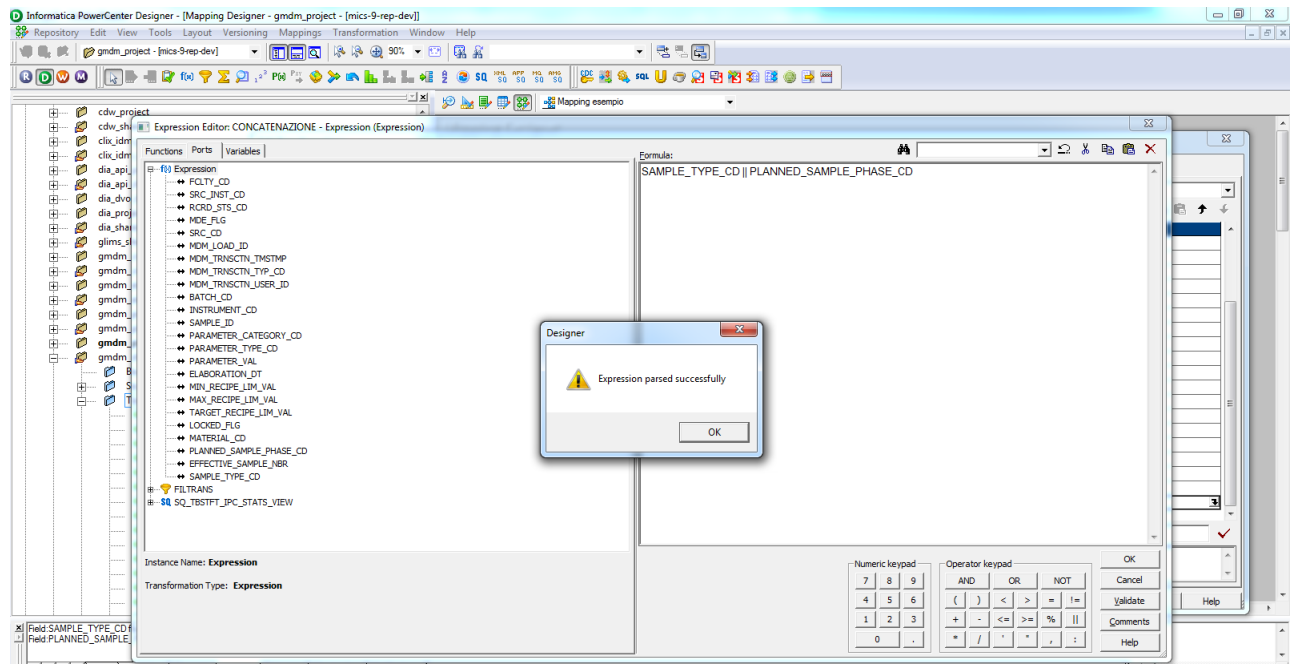


Figura 6.8: Oggetto *expression* del designer

Due oggetti di trasformazione, che meritano di essere attenzionati in quanto usati in tutti i *mapping* creati durante il lavoro svolto, sono l'oggetto *lookup* e l'oggetto *update strategy*. Gli oggetti indicati, vengono usati in combinazione per andare a definire la politica di inserimento dei dati sulla tabella *target*. Tramite essi è possibile andare a verificare la presenza o meno degli attributi chiave che si tenta di inserire nella *target*, se non è presente si andrà a inserire l'intera riga, mentre se la chiave è già presente in *target* si andrà ad aggiornare gli attributi non chiave della corrispondente tupla.

Per prima cosa si andrà a definire l'oggetto *lookup*. Questo oggetto di trasformazione di tipo attivo/passivo può, a differenza degli altri oggetti di trasformazione indicati precedentemente, essere sia scollegato che collegato al flusso dati.

Una trasformazione sconnessa non sarà collegata ad altre trasformazioni del *mapping*, ma sarà chiamato in un'altra trasformazione, restituendo un valore per quella trasformazione. Tramite questa trasformazione, è possibile cercare dati in un *file flat*, tabella relazionale o da qualsiasi altra fonte per cui sia PowerCenter sia *integration service* possono connettersi.

L' *integration service*, interroga la fonte di ricerca in base alle porte di ricerca nella trasformazione e di una condizione di ricerca. L'oggetto *lookup* restituirà il risultato della ricerca per la *target* o per un'altra trasformazione. Le specifiche d'uso di questi due oggetti, verranno ripresi nel Capitolo 7.

#### 6.4.4 Workflow manager

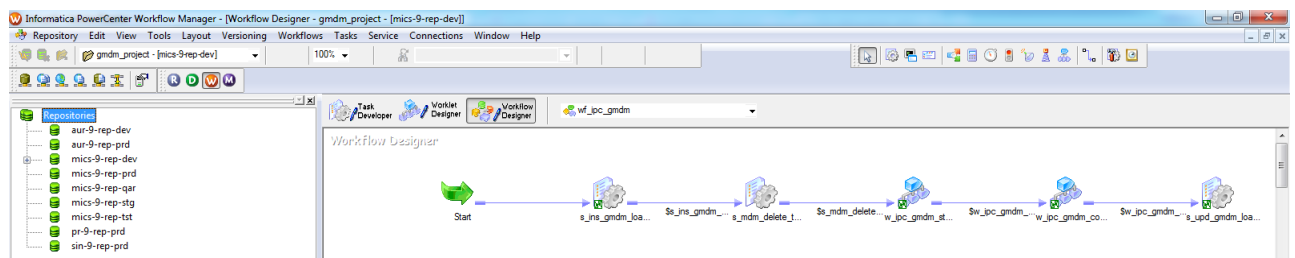
Nel *workflow manager*, si definisce un insieme di istruzioni chiamato flusso di lavoro (*work flow*), per eseguire i *mapping* definiti con il designer.

In generale, un "work flow" contiene una sessione e qualsiasi altra attività (*task*) che può essere eseguita quando si esegue una sessione.

I "task" possono includere sessioni, notifiche *e-mail*, o informazioni di schedulazione. Ogni *task* è collegato nel *work flow* tramite *link*, specificando l'ordine di esecuzione dei *task* creati. Tramite il *workflow manager* è possibile anche definire le *worklet*. Una *worklet* è un oggetto che raggruppa un insieme di *task*, è simile ad un *work flow*, ma senza informazioni di pianificazione. È possibile eseguire una serie di *worklet* all'interno di un *work flow*.

Il *workflow manager* (Fig. 6.9) è costituito da tre strumenti per assistere nella creazione dei *work flow*:

- il *task developer*, utilizzato per creare i *task* che si desidera eseguire nel *work flow*;
- il *workflow designer*, utilizzando per creare un *work flow*, andando a specificare l'ordine di esecuzione dei *task* mediante *link*;
- il *worklet designer*, utilizzato per creare le *worklet*.



**Figura 6.9:** Finestra principale del *workflow manager*

In generale tramite il *task developer* è possibile definire i seguenti tipi di *task*:

- *assignment*, utilizzato per assegnare un valore a una variabile del *work flow*;
- *command*, per specificare un comando di *shell* da eseguire durante il *work flow*;
- *control*, per interrompere o abortire il *work flow*;
- *decision*, per definire una condizione che determina l'esecuzione del *work flow*;
- *e-mail*, per inviare *e-mail* durante l'esecuzione del *work flow*;
- *event-raise*, utilizzato per attivare un evento;
- *event-wait*, per attendere che si verifichi un evento, prima di eseguire il *task* successivo;
- *session*, per eseguire i *mapping* creati con il *designer*;
- *timer*, per permettere all' *integration service* di eseguire il *task* ad un determinato orario o dopo un periodo di tempo prestabilito.

Per i *work flow* trattati in questo elaborato, i *task* più usati sono naturalmente le *session*. Una *session* è un insieme di istruzioni che indicano all' *integration service* come e quando spostare i dati dalla sorgente alla *target*. Infatti durante la creazione di una *session* si deve specificare il *mapping* che si desidera eseguire dall' *integration service*. Quindi ad ogni *session task* corrisponderà un *mapping* creato nel *designer*.

Quando si crea una *session task*, è possibile settare informazioni, come il nome della *session*, l' *integration service* per eseguire la *session*, selezionare le opzioni per eseguire comandi di *shell* pre-session, settare l'invio di *e-mail* di tipo *on-success* o *on-failure*, specificare le connessioni degli oggetti del *mapping*, come le sorgenti, i *target* o le *lookup*.

Una volta definiti i *task* e il *work flow* che conterrà i *task* creati, tramite il *workflow manager* è possibile settare la schedulazione. Una schedulazione è un oggetto di *repository* che conterrà una serie di impostazioni di pianificazione. Ciascun *work flow* avrà uno schedatore associato. Tramite la schedulazione è possibile pianificare l'esecuzione continua di un *work flow*, o la ripetizione in un determinato momento o intervallo (6.10).

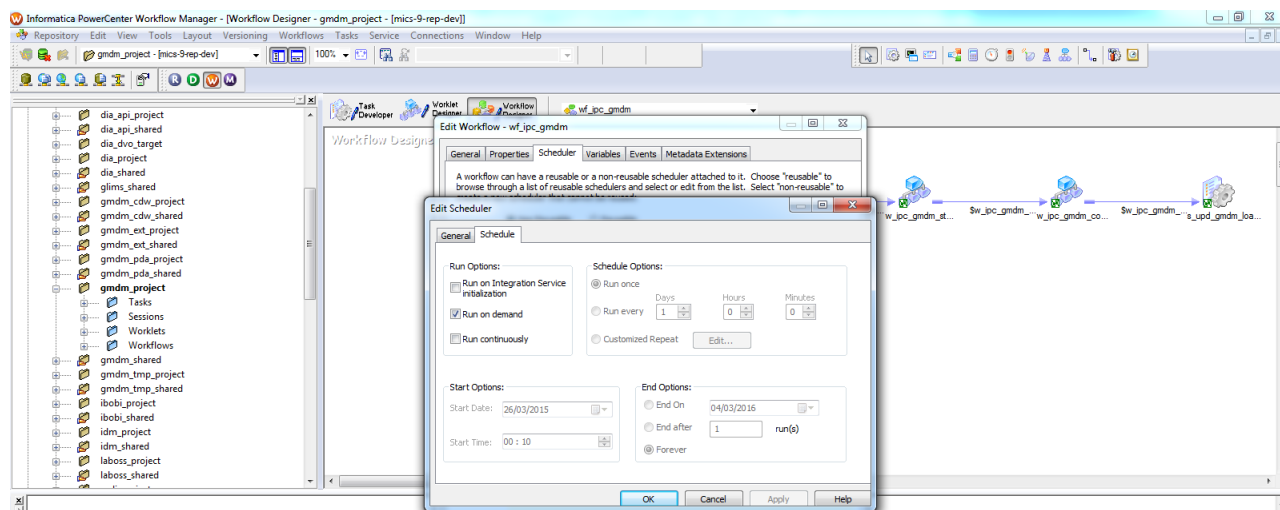


Figura 6.10: Schedatore del *workflow manager*

Di default, lo schedatore è impostato su "run on demand" (in questo caso l'*integration service* esegue il *work flow*, quando sarà avviato manualmente dall'utente). Se un *work flow* si arresta per errore, la schedulazione si disabiliterà automaticamente. In questo caso sarà necessario riattivare manualmente la schedulazione.

### 6.4.5 Workflow monitor

Tramite il *workflow monitor* è possibile monitorare cronologicamente e in tempo reale i dettagli relativi all'esecuzione di *work flow*, che fanno parte del *repository* su cui si è connessi (Fig. 6.11).

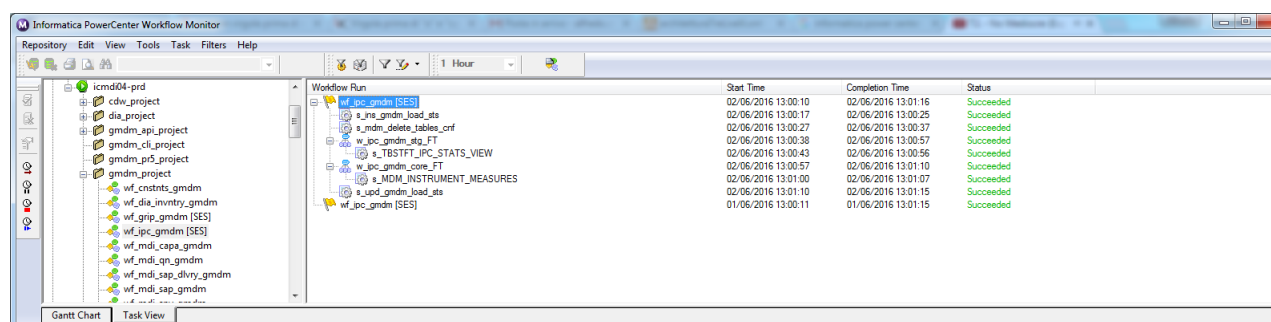
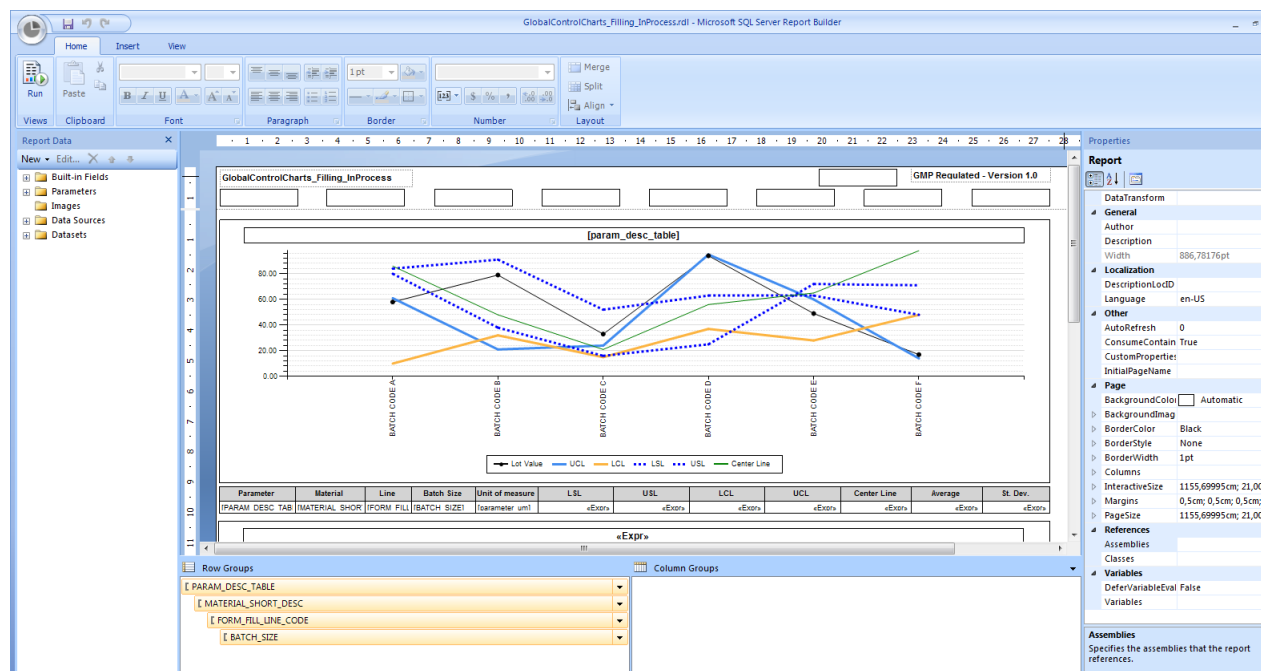


Figura 6.11: Finestra principale del *workflow monitor*

Il *workflow monitor* permette di avviare, fermare o bloccare l'esecuzione dei flussi, visualizzare statistiche relative agli oggetti monitorati, analizzare nome, tempo di inizio e stato dei flussi, visualizzare il numero delle righe che sono state caricate con successo nei *target* e visualizzare la schedulazione dei flussi. In caso di errori di esecuzione o per visualizzare dettagli più specifici è possibile visualizzare il file di log.

## 6.5 Microsoft SQL Server Reporting Services Report Builder

"Microsoft SQL Server Reporting Services Report Builder" (Fig. 6.12) è lo strumento utilizzato nel corso del lavoro per realizzare i *report* di analisi, destinati ad essere distribuiti all'interno dell'organizzazione attraverso la condivisione sulla piattaforma di "Microsoft Share Point".



**Figura 6.12:** Finestra principale di Microsoft SQL Server Reporting Services Report Builder

Con "Report Builder" è possibile presentare dati analitici graficamente, progettando *report* che possono essere visualizzati all'interno di pagine *web*, in formato PDF, oppure esportati in formati accessibili da excel.

Al giorno d'oggi lo sviluppo di un *report* comprende una gamma molto ampia di requisiti, ed è spesso difficile mettere un limite a quello che dovrebbe essere previsto da un sistema di reportistica. Per questo motivo, i progettisti di Report Builder hanno definito questo ampio insieme di requisiti nella seguente equazione:  $data\ access + data\ process + report\ layout = report\ output$ .

"Data access", è il recupero dei dati da una tabella di un *database*, un file di testo, un oggetto *delphi*, o qualsiasi altro *dataset* con struttura tabellare.

"Data process", si riferisce al processo di calcolo dei risultati intermedi sulla base di dati.

"Report layout", è un insieme di componenti che descrivono l'aspetto del documento e definiscono il comportamento di essi durante la generazione del *report*.

"Report output", è un insieme di componenti che descrivono l'esatto contenuto di ogni singola pagina.

La creazione di un *report* con Report Builder può essere suddiviso in quattro attività principali:

- *select*, si riferisce alla selezione di dati;
- *design*, si riferisce alla creazione di un *layout* che descrive come il documento dovrebbe apparire all'utente che ne farà uso;
- *process*, si riferisce alla manipolazione dei dati o del *layout* per controllare la generazione del documento;

- *Generate*, si riferisce alla creazione del documento effettivo.

Al fine di selezionare i dati che andranno a popolare il *report*, è necessario creare un'origine dati.

In Report Builder ci sono due tipi di origine dati:

- *integrata*, in questo caso è necessario disporre delle autorizzazioni e delle informazioni per la connessione a una fonte esterna di dati necessaria per accedere ai dati;
- *condivisa*, in questo caso è possibile selezionare l'origine dati condivisa in un *server* di Microsoft Share Point.

Successivamente si creano uno o più insieme dati per ogni fonte. Ogni *set* di dati specifica i campi della sorgente dati che si prevede di utilizzare nel *report*. Questa fase permette oltre a selezionare i dati, a manipolarli al fine di aumentare il significato del dato per il *report*. Questa attività viene eseguita tramite lo strumento "*query designer*" che semplifica notevolmente la selezione dei dati senza richiedere una conoscenza approfondita sulla basi di dati.

Per permettere il controllo del *report* è necessario definire i parametri.

In Report Builder i "parametri" possono essere utilizzati in due diversi modi: per filtrare i dati nei *dataset* creati o per filtrare i dati nel *report*. La restrizione dei dati nei *dataset* permettono di migliorare le prestazioni di visualizzazione di un *report*. È anche possibile utilizzare i parametri per ordinare e organizzare i dati in un *report*.





## Capitolo 7

# ESTRAZIONE, TRASFORMAZIONE E CARICAMENTO

Si descrive nel seguito il processo di ETL che include tutte le operazioni realizzate e le procedure implementate per integrare la sorgenti di dato disponibili e caricare le informazioni richieste all'interno del *data mart* progettato. Il processo globale include le seguenti macro fasi, ognuna con un obiettivo specifico:

- Integrazione di *IPC Book* nel GMDM: che si concretizza con il caricamento dell'area di *staging* e del *core* con la nuova sorgente.
- Integrazione del *data mart*: che si concretizza con il caricamento dell'area dell'*extension*, con le informazioni presenti nel *core*, in seguito all'integrazione eseguita nella fase precedente.

Nel nostro lavoro tale fase insieme alla creazione del *report* è stato quello che ha comportato il maggior impiego di tempo e risorse, circa il 70% del totale. Sono stati implementati due flussi d'integrazione dei dati di tipi *batch-oriented*, cioè schedulati per essere eseguiti ad intervalli temporali prestabiliti.

In questo capitolo si illustreranno le modalità con le quali è stato impostato il processo di ETL e i flussi che sono stati realizzati nel corso del progetto.

I *work flow* sviluppati in questo lavoro sono:

- Il flusso **wf\_ipc\_gmdm**: che permette l'integrazione della sorgente *IPC Book* nell'architettura GMDM. Tramite questo flusso verranno caricati sia i due livelli di *staging* che quello del *core*.
- Il flusso **wf\_filling\_control\_chart**: che permette l'integrazione del *data mart*, presente nell'area dell'*extension*, a partire dalla nuova integrazione avvenuta nel *core*.

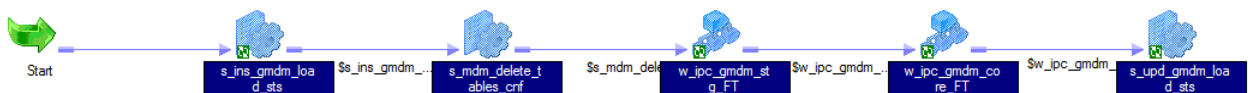
Per questo lavoro, è stato integrato *IPC Book* con i dati provenienti dal sito produttivo di Sesto Fiorentino. Ma peculiarità del GMDM (Sez. 3.2) è quella di essere una architettura globale, quindi dovrà essere possibile, usare gli stessi flussi, per l'integrazione della stessa sorgente per siti produttivi differenti. Questo è possibile, tramite:

- L'uso di un file dei parametri, uno per flusso, andando a specificare, oltre alle connessioni, gli attributi tecnici (FCLTY\_CD e SRC\_CD) specifici per il caricamento.
- Qualsiasi filtro o controllo nei *mapping* o *session* sulla *facility* e sulla sorgente, dovrà essere fatto tramite variabili valorizzate nel file dei parametri.

In questo modo, i *work flow* creati, avranno per *input* un file dei parametri specifico per sito produttivo. Così facendo, quando verrà richiesto di integrare nell'architettura, la sorgente IPC *Book* per un altro sito produttivo, basterà creare un nuovo file dei parametri per il flusso "wf\_ipc\_gmdm".

In questo capitolo si illustreranno le modalità con le quali è stato impostato il processo di ETL e le caratteristiche dei flussi che sono stati realizzati nel corso del progetto.

## 7.1 Integrazione dell'area di *staging* e del *core*



**Figura 7.1:** Work flow per l'integrazione dell'area di *staging* e del *core*

Il flusso "wf\_ipc\_gmdm" (Fig. 7.1), contiene:

- Le session "s\_ins\_gmdm\_load\_sts" e "s\_upd\_gmdm\_load\_sts" di tipo *reusable*, usate in tutti i flussi del *core* del GMDM. Queste due session, permettono di aggiornare le seguenti tabelle di sistema del GMDM:
  - MDM\_LOAD\_STS: che contiene il *timestamp* di inizio e fine, lo stato di completamento (*succeeded/error*) e l'identificativo univoco di caricamento, per ogni flusso di integrazione nel *core* del GMDM.
  - GMDM\_STG\_LOAD\_STS: che contiene per ogni flusso del *core* del GMDM l'ultima valorizzazione delle finestre temporali usate per il caricamento.
- La session "m\_mdms\_delete\_tables\_cnf" di tipo non *reusable*, usata in tutti i flussi del *core* del GMDM. Questa session, permette di aggiornare le seguenti tabelle di sistema del GMDM:
  - MDM\_DELETE\_TABLES\_CNF: che contiene per ogni tabella del *core* e dell'*extension* del GMDM l'ultima modalità di caricamento (*full/delta*) con il massimo *timestamp* presente nella tabella *target*. Questa tabella viene utilizzata per le procedure di *delete* dei *record* non aggiornati delle tabelle del GMDM.
  - MDM\_STG\_LOAD\_STS: che contiene per ogni flusso del *core* del GMDM l'ultima valorizzazione delle finestre temporali usate per il caricamento, con la modalità di caricamento (*full/delta*). Questa tabella viene usata durante il caricamento delle tabelle del *core* per permettere di filtrare i *record* della sorgente sul *timestamp*.
- La *worklet* "w\_ipc\_gmdm\_stg\_FT" di tipo *reusable* che contiene la session per il caricamento dell'area di *staging*.
- La *worklet* "w\_ipc\_gmdm\_core\_FT" di tipo *reusable* che contiene la session per il caricamento dell'area del *core*.

Per permettere di aggiornare nella tabella "MDM\_LOAD\_STS", lo stato finale di esecuzione del *work flow*, ogni nodo presente nel flusso, passa al successivo il suo stato finale. Se un nodo fallisce, non verranno eseguiti tutti i nodi successivi e l'intero *work flow* andrebbe in errore.

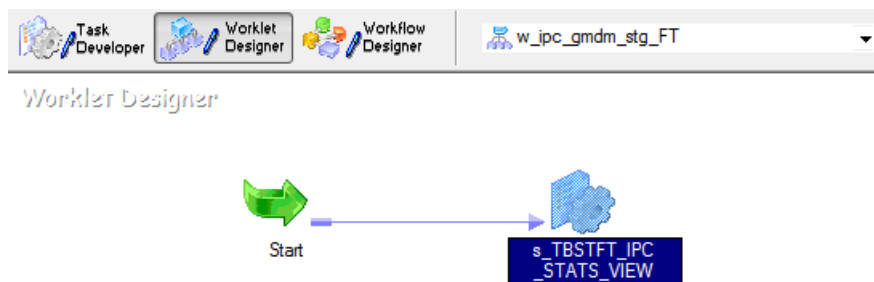
### 7.1.1 Integrazione dell'area di staging

Come descritto nella sezione 3.2, l'area di staging, viene caricata tramite la *store procedure* parametrica "LOAD\_STG". La *store procedure* integra le tabelle di *staging* utilizzando la tabella di configurazione "TBSTCNF\_LOAD\_STG", dove vengono definiti i parametri necessari per il caricamento dei due livelli di *staging* del GMDM (Sorgente A.1).

Gli attributi più importanti definiti nella tabella TBSTCNF\_LOAD\_STG sono:

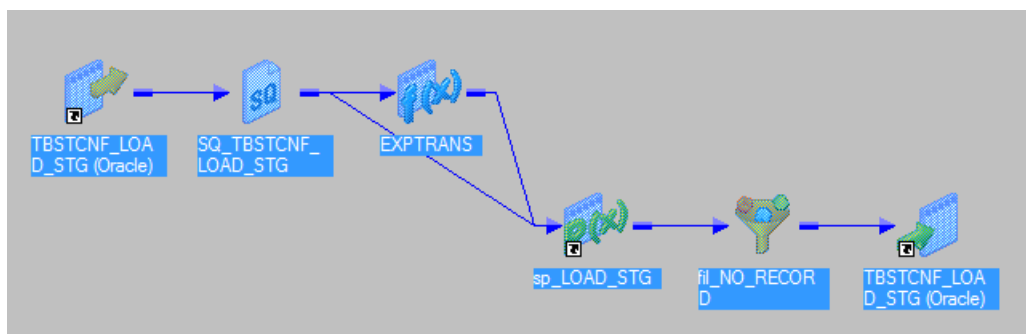
- SRC\_SYSTEM\_CD: identificativo della sorgente dalla quale si vanno a reperire i dati;
- SRC\_TABLE\_NAME\_CD: identificativo del nome della tabella nel sistema sorgente dalla quale bisogna estrarre i dati;
- SRC\_FACILITY\_CD: identificativo sito produttivo;
- SRC\_INST\_CD: identificativo dell'installazione sul sito produttivo;
- SRC\_DBLINK\_CD: all'interno di questo campo viene inserito il nome del *database link* da utilizzare per reperire i dati dalla sorgente;
- STG\_SCHEMA\_1\_LEV: è lo schema da utilizzare per le tabelle di *staging* di primo livello;
- STG\_SCHEMA\_2\_LEV: è lo schema da utilizzare per le tabelle di *staging* di secondo livello;
- STG\_TABLE\_NAME\_CD: indentificativo del nome della tabella *target* di secondo livello;
- SRC\_FILTER\_DESC: per specificare eventuali filtri da utilizzare sulla tabella sorgente;
- SRC\_SCHEMA: è lo schema del sistema sorgente.

La *store procedure* viene invocata nel mapping "m\_STG\_LOAD" associato alla *session* "s\_TBSTFT\_IPC\_STATS\_VIEW", contenuta nella *worklet* "w\_ipc\_gmdm\_stg\_FT" (Fig. 7.2).



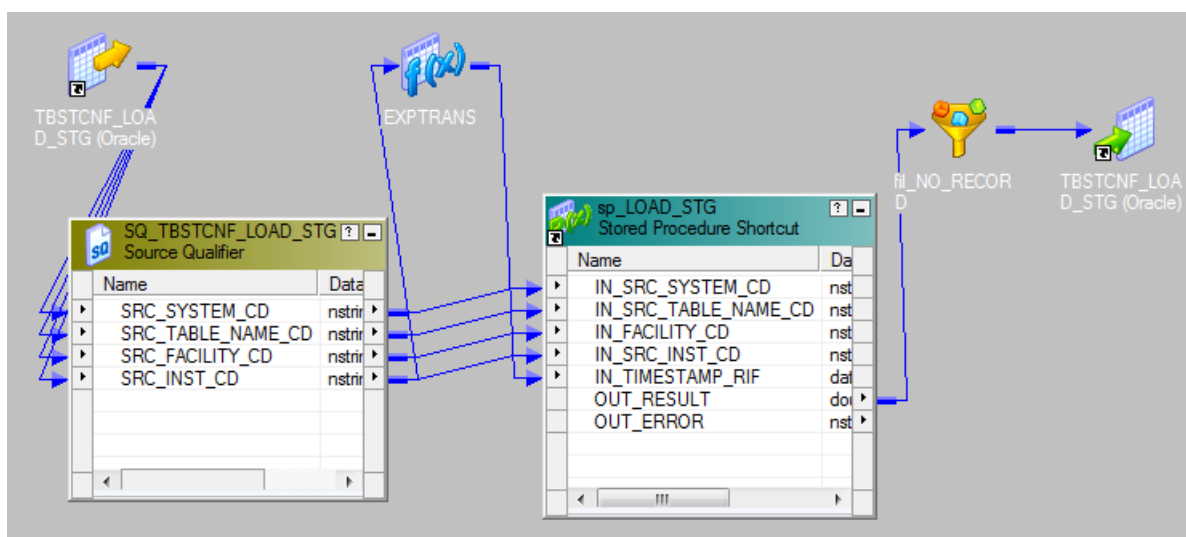
**Figura 7.2:** Worklet w\_ipc\_gmdm\_stg\_FT contenuta nel work flow wf\_ipc\_gmdm

Il *mapping* "m\_STG\_LOAD" ha come sorgente la tabella di configurazione e l'oggetto "sp\_LOAD\_STG" per eseguire la procedura (Fig. 7.3).



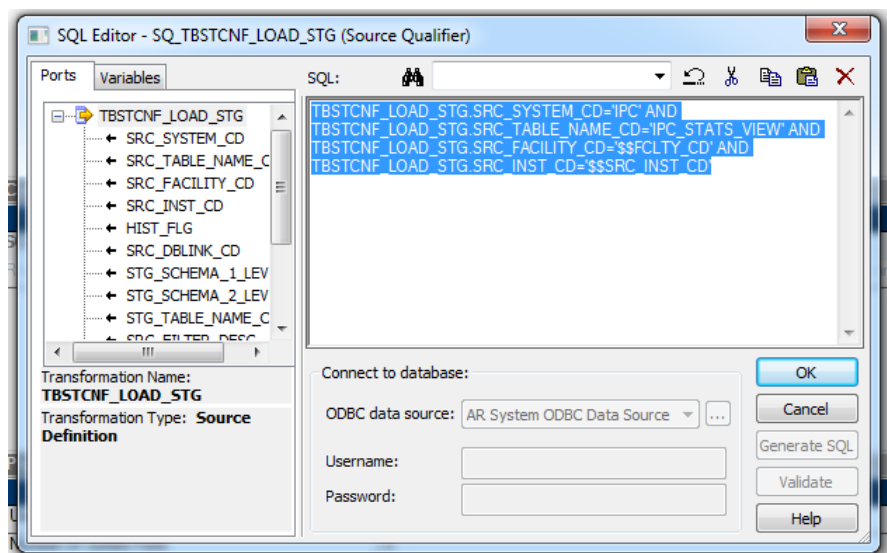
**Figura 7.3:** Mapping m\_STG\_LOAD associato alla session s\_TBSTFT\_IPC\_STATS\_VIEW

Il passaggio dei parametri alla procedura, avviene tramite l'oggetto "sp\_LOAD\_STG" (Fig. 7.4).



**Figura 7.4:** Oggetto sp\_LOAD\_STG per eseguire la store procedure LOAD\_STG

Dato che il *mapping* m\_STG\_LOAD è unico per tutti i *work flow* del *core*, la restrizione sulla tabella sorgente, per selezione i campi tecnici di IPC Book è definita nella *session* s\_TBSTFT\_IPC\_STATS\_VIEW (Fig. 7.5).

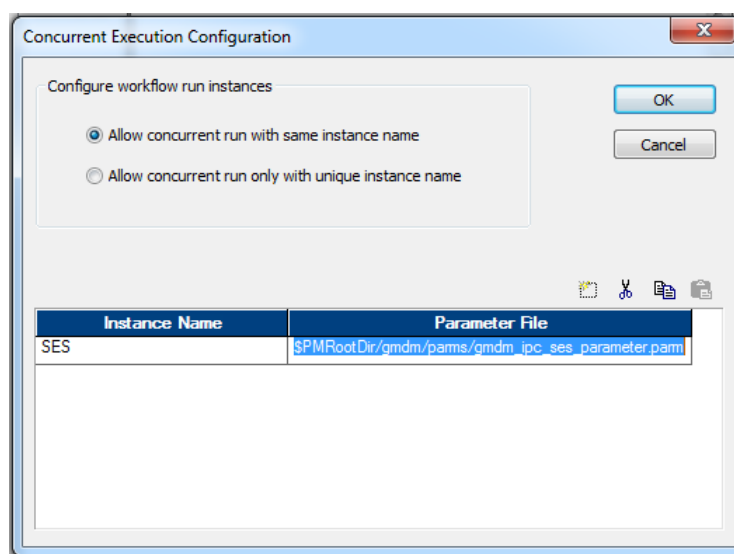


**Figura 7.5:** Restrizione sulla tabella di configurazione TBSTCNF\_LOAD\_STG

Come si può notare in figura 7.5, la restrizione sui campi tecnici "FCLTY\_CD" e "SRC\_INST\_CD" viene effettuata con l'uso delle variabili.

Le variabili che contengono il prefisso "\$\$", vengono dichiarate dove servono (*mapping*, *session*, *worklet* o *work flow*) e la loro valorizzazione sarà contenuta nel file dei parametri passato in ingresso al *work flow* (Sorgente B.1).

È possibile così definire, esecuzioni concorrenti sullo stesso *work flow* (Fig. 7.6).

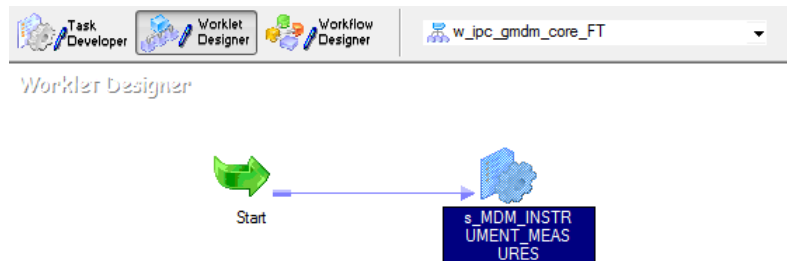


**Figura 7.6:** Configurazione di esecuzioni concorrenti

L'oggetto *target* contenuto nel *mapping* di figura 7.3 è fittizio, in quanto serve per rendere il flusso valido.

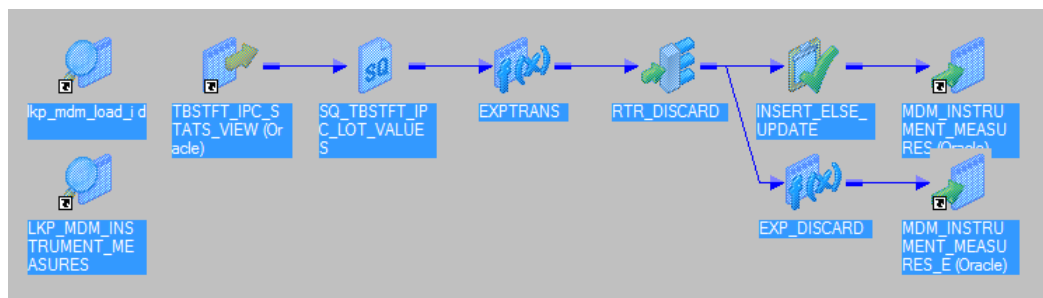
### 7.1.2 Integrazione dell'area del *core*

I dati provenienti da IPC *Book* integrati nello *staging* serviranno a caricare la tabella "MDM\_INSTRUMENT\_MEASURES" del *core*. Il caricamento viene effettuato dalla *session* "s\_MDM\_INSTRUMENT\_MEASURES", contenuta nella *worklet* "w\_ipc\_gmdm\_core\_FT" (Fig. 7.7).



**Figura 7.7:** Worklet w\_ipc\_gmdm\_core\_FT contenuta nel workflow wf\_ipc\_gmdm

Nell'immagine 7.8 è possibile vedere il *mapping* associato alla *session* "s\_MDM\_INSTRUMENT\_MEASURES".



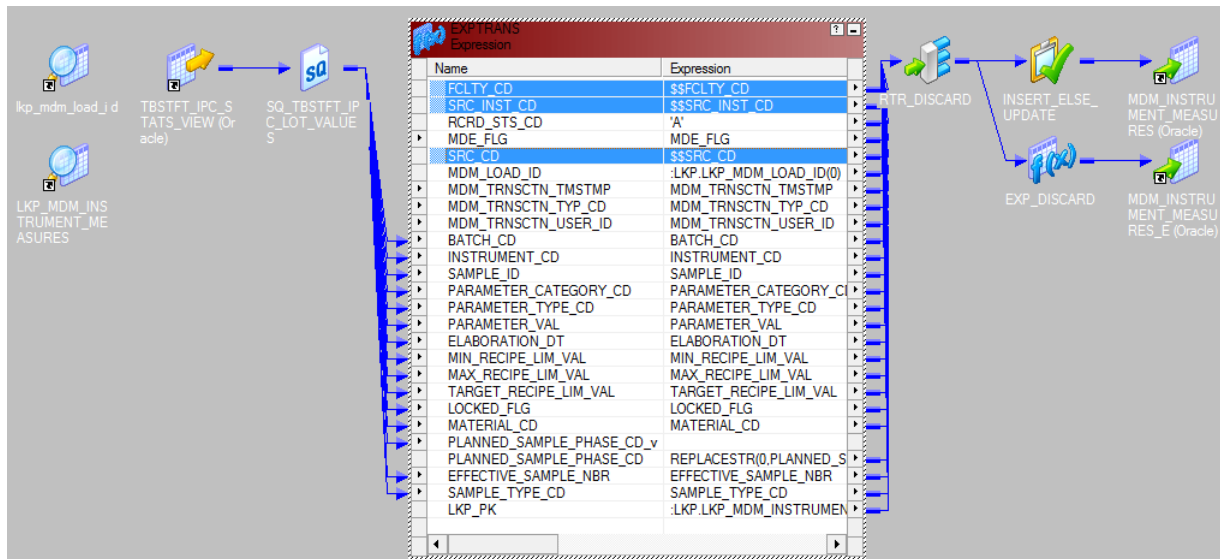
**Figura 7.8:** Mapping per l'integrazione dell'area del *core*

In questo caso le trasformazioni riguardano principalmente i campi della tabella *target*.

La sorgente è la sola tabella "TBSTFT\_IPC\_STATS\_VIEW" presente nello *staging* di secondo livello del GMDM.

Nel *source qualifier* vengono selezionati i campi di interesse filtrando i *record* sul *timestamp* valorizzato nella tabella "MDM\_STG\_LOAD\_STS" (Sorgente A.2).

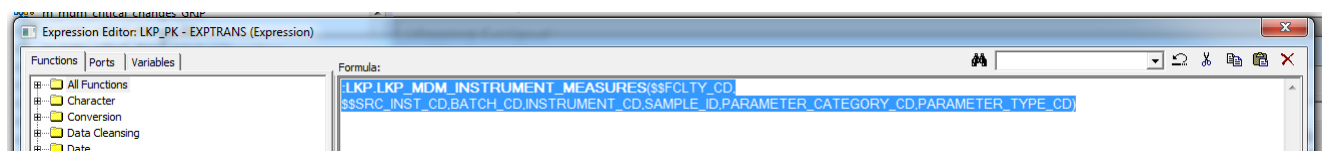
Nell'oggetto "expression" vengono effettuate le modifiche sui campi della tabella *target*. Gli attributi tecnici della tabella *target* vengono valorizzati con le variabili presenti nel *file dei parametri* (Fig. 7.9).



**Figura 7.9:** Oggetto *expression* definito nel *mapping* per l'integrazione dell'area del *core*

Il campo "MDM\_LOAD\_ID" (Fig. 7.9), viene valorizzato tramite la chiamata alla *lookup* scollegata "lkp\_mdm\_load\_id". La *lookup* seleziona l'ultimo identificativo dell'esecuzione del *work flow* aggiornato dalla *session* "s\_ins\_gmdm\_load\_sts".

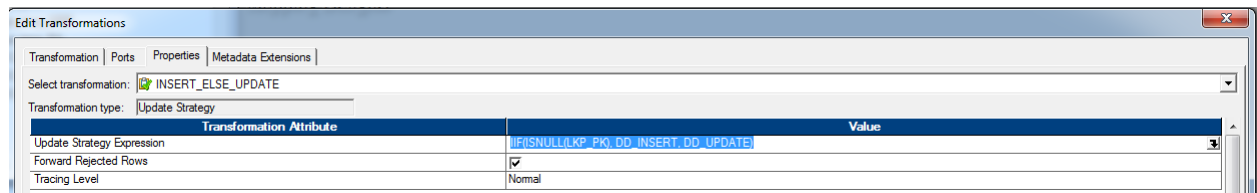
Nell'*expression* è presente un'altra chiamata alla *lookup* scollegata "LKP\_MDM\_INSTRUMENT\_MEASURES". Questa *lookup* serve a determinare se il caricamento sulla *target* deve essere un "insert" o un "update". Nella *lookup* vengono passati gli attributi provenienti dal flusso, che formano la chiave primaria della tabella *target* (Fig. 7.10).



**Figura 7.10:** Oggetto *expression* per la strategia di aggiornamento della tabella *target*

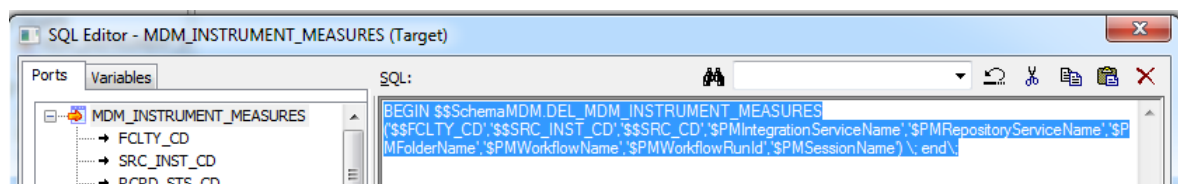
L'oggetto *lookup*, andrà a confrontare ogni valore in ingresso con la corrispondente chiave della tabella *target*. L'*integration service* interrogherà la fonte della *lookup* e restituirà un valore nella porta di ritorno, che sarà passato nell'espressione ":LKP". Nel nostro caso, l'oggetto di ritorno sarà "null", quando non troverà la corrispondenza della chiave e diverso da "null" quando invece la chiave sarà già presente nella *target*.

Non resta che controllare l'inserimento o l'aggiornamento sulla *target* tramite l'oggetto *update strategy*. Quando il valore sarà *null* si andrà a inserire la nuova riga sulla *target*, mentre si aggiornerà la riga sulla *target* nel caso contrario (Fig. 7.11).



**Figura 7.11:** Oggetto *update strategy* per la strategia di aggiornamento della tabella *target*

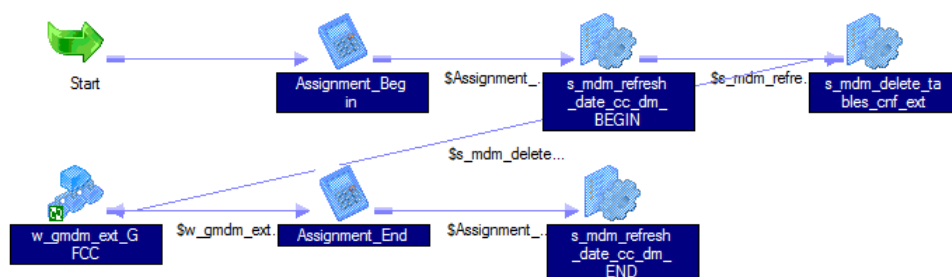
La procedura di "delete" viene richiamata come *post SQL* nella *session* *s\_MDM\_INSTRUMENT\_MEASURES* (Fig. 7.12).



**Figura 7.12:** Chiamata alla procedura di *delete* per la tabella del *core*

La procedura (Sorgente A.4) elimina dalla *target* tutti i *record* con *timestamp* minore rispetto all'ultimo *timestamp* del caricamento in *full* presente nella tabella "MDM\_DELETE\_TABLES\_CNLF. I record eliminati verranno inseriti nella corrispondente tabella degli scarti.

## 7.2 Integrazione dell'area dell'*extension*



**Figura 7.13:** *Work flow* per l'integrazione dell'area dell'*extension*

Il flusso "wf\_filling\_control\_chart" (Fig. 7.13), contiene:

- Due *task* "Assignment\_Begin" e "Assignment\_End" per valorizzare la variabile "\$\$Wf\_Begin\_End\_flg". All'inizio la variabile, sarà valorizzata con "B". Se ogni *task* del *work flow* verrà eseguito correttamente la variabile tramite il nodo "Assignment\_End" sarà valorizzata con "E".
- Le *session* "s\_mdm\_refresh\_date\_cc\_dm\_BEGIN" e "s\_mdm\_refresh\_date\_cc\_dm\_END" di tipo non *resuable*, usate per aggiornare, sulla base del valore assunto dalla variabile "\$\$Wf\_Begin\_End\_flg", la tabella "MDM\_REFRESH\_DATE\_CC\_DM". Questa tabella contiene il *timestamp* di inizio e fine, lo stato (*succeeded/error*) e l'identificativo univoco di caricamento del *work flow*.



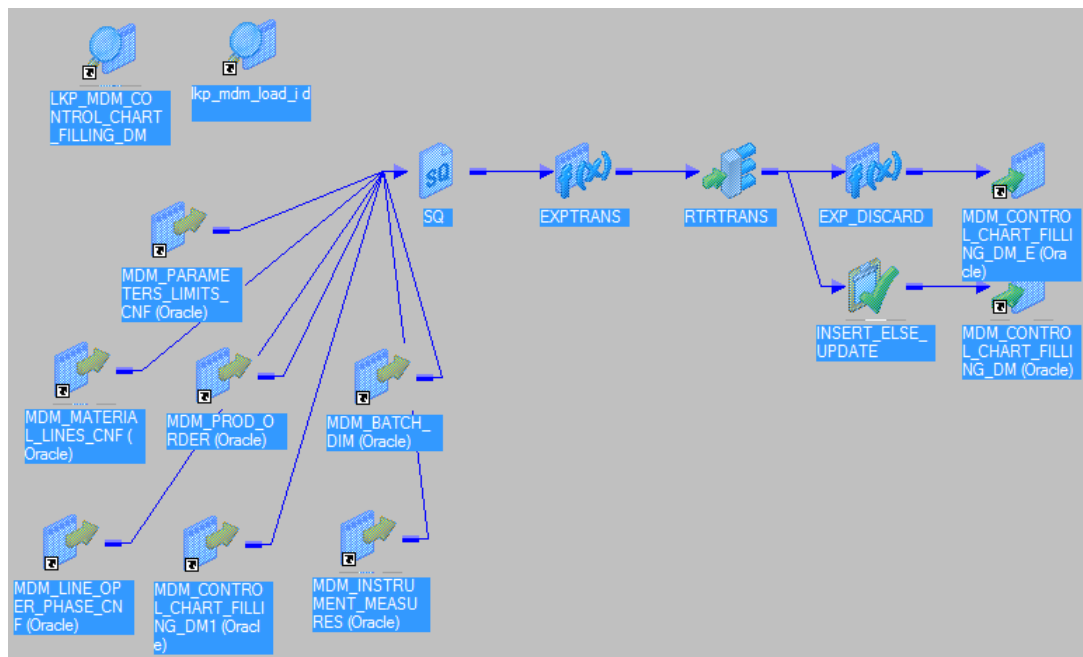
- La session "s\_mdm\_delete\_tables\_cnf\_ext di tipo non *reusable*, simile a quella presente nel *work flow* di integrazione del *core*. Permette di aggiornare la tabella "MDM\_DELETE\_TABLES\_CNF" con il massimo *timestamp* presente nella tabella *target*. Anche in questo caso, la tabella sarà utilizzata per le procedura di *delete* dei *record* non aggiornati sul *data mart*.
- La *worklet* "w\_gmdm\_ext\_GFCC" di tipo "*reusable*", che contiene la *session* per il caricamento del *data mart*.

Il caricamento viene effettuato dalla *session* "s\_mdm\_control\_chart\_filling\_dm", contenuta nella *worklet* "w\_gmdm\_ext\_GFCC" (Fig.7.14).



**Figura 7.14:** Worklet w\_gmdm\_ext\_GFCC contenuta nel *work flow* wf\_filling\_control\_chart

In figura 7.15 è possibile visualizzare il *mapping* associato alla *session* "s\_mdm\_control\_chart\_filling\_dm".

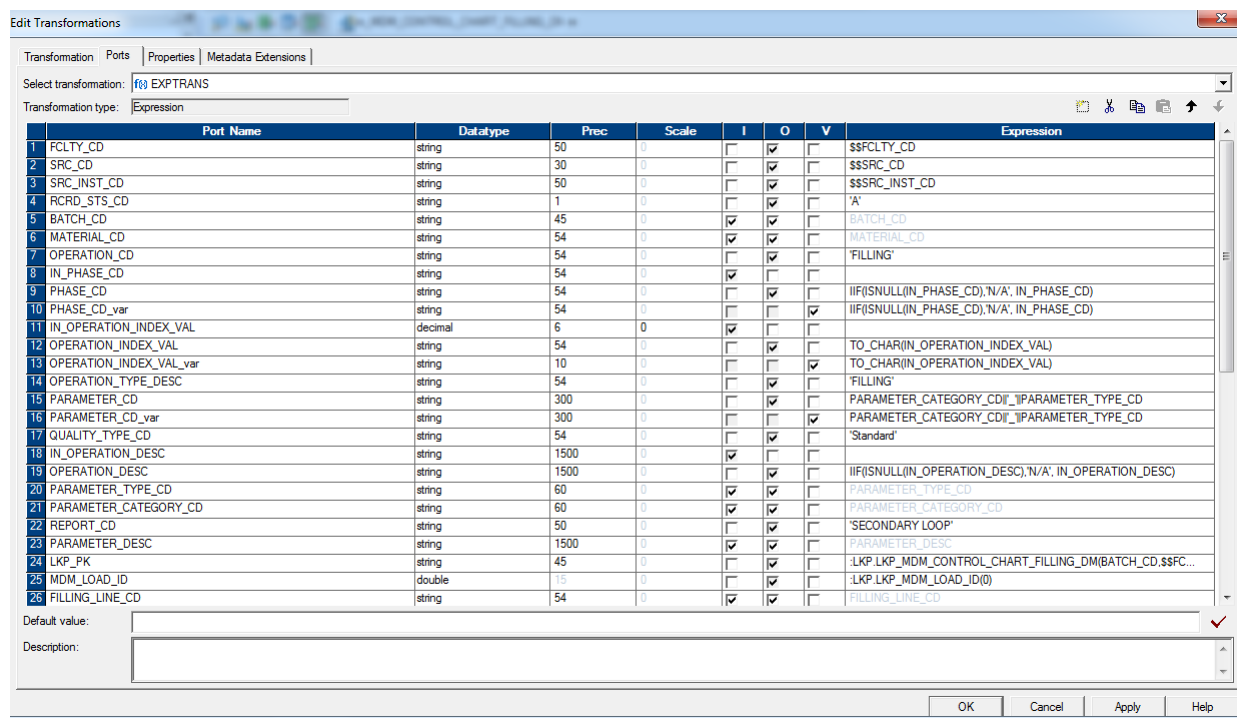


**Figura 7.15:** Mapping per l'integrazione dell'area dell'extension

Come si nota dall'immagine, il *mapping* contiene due *lookup* scollegate ("LKP\_MDM\_CONTROL\_CHART\_FILLING\_DM" e "lkp\_mdm\_load\_id"), con la stessa funzionalità del flusso "wf\_ipc\_gmdm".

Le sorgenti del *data mart*, (come analizzato nella sezione 5.3) non sarà la sola tabella del *core* "MDM\_INSTRUMENT\_MEASURES". Le logiche di *join* e di interrogazione delle sorgenti vengono definite anche in questo caso nel "*source qualifier*" (Sorgente A.3).

Nel *mapping*, non vengono effettuate trasformazioni di tipo "attive". L'uniche trasformazioni effettuate, sono di tipo "passive" e sono riportate nell'immagine 7.16.



**Figura 7.16:** Oggetto *expression* definito nel *mapping* per l'integrazione dell'area dell'*extension*

Gli attributi tecnici del *data mart* vengono aggiornati con le variabili valorizzate nel corrispondente *file* dei parametri (Sorgente B.2).

Anche per l'*extension*, la procedura di "delete" viene richiamata come *post SQL* nella *session* "s\_mdm\_control\_chart\_filling\_dm" (Sorgente A.5).

### 7.3 Schedulazione dei *work flow*

Come spiegato nella sezione 6.4.4, la schedulazione definita con il *workflow manager*, ha il difetto di disabilitarsi automaticamente qualora un flusso si arresti per errore. Per questo motivo, la società ha deciso di ricorrere al *crontab* di *unix* (Sorgente C.1).

I flussi saranno pianificati per girare una volta al giorno, inoltre come facilmente ci si aspetta il flusso dell' *extension* girerà dopo l'esecuzione del *core*.

## Capitolo 8

# INTERFACCIA UTENTE E REPORTISTICA

Il capitolo che segue ha come obiettivo quello di mostrare le ultime fasi di realizzazione della soluzione "*global filling control charts*". In particolare, terminata la progettazione del *data mart* e implementati i processi di ETL volti a caricare i dati nelle strutture definite, si presenta la creazione dell'interfaccia rivolta agli utenti del sistema.

### 8.1 Fasi principali per la realizzazione

Per questo elaborato è stato realizzato un report realizzato con "Report Builder", dove grazie a tabelle, istogrammi e carte di controllo l'utente può effettuare analisi di *trend* dei parametri di interesse durante la fase di "*filling*".

Il report è formato da:

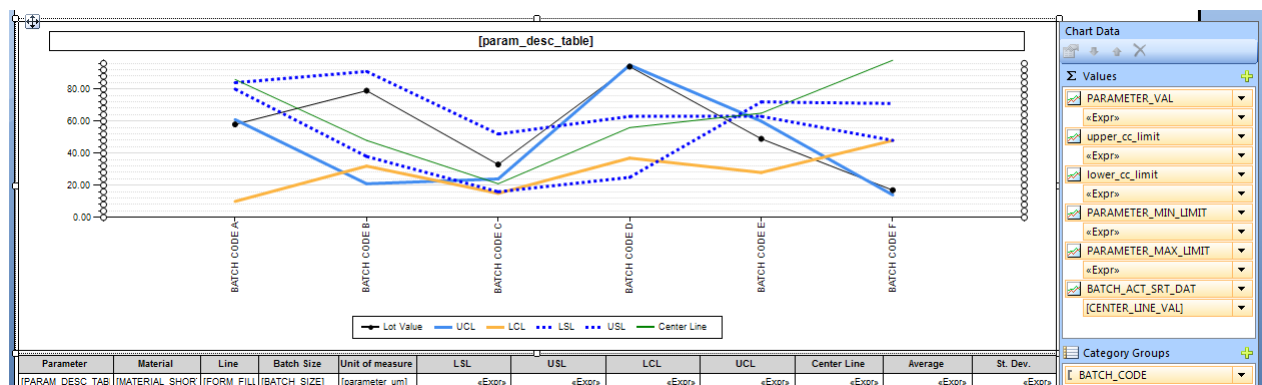
- Un menù superiore dove è possibile selezionare il parametro di analisi.
- Una prima carta di controllo contenente per ogni *batch* o *sample* il valore del parametro selezionato con i rispettivi limiti.
- Una tabella per riassumere il risultato ottenuto nella prima carta di controllo. La tabella conterrà per ogni parametro, materiale del *batch*, linea di riempimento del *batch* e dimensione del *batch*, i limiti, la media e la deviazione *standard*.
- Una seconda carta di controllo contenente per ogni *batch* o *sample* il valore del limite di specificazione inferiore (LSL) e del limite di specificazione superiore (USL).
- Un istogramma con i valori della capacità di processo ( $C_{pk}$ ) per *batch* o *sample*.

Il *report* è sostanzialmente statico ma grazie all'impostazione dei *prompt* (posto a destra della finestra di visualizzazione), conferisce una certa dinamicità, consentendo agli utenti di personalizzarlo in base alle proprie esigenze.

Caratteristica fondamentale del *report* sono i limiti, questi come abbiamo visto nella sezione 5.3, non sono stati integrati nel *data mart* ma su una tabella di configurazione in modo da essere facilmente modificabili su richiesta dell'utente. Questo è stato reso possibile grazie alla possibilità di Report Builder di definire al suo interno dei *data set*, dove è possibile integrare del linguaggio SQL.

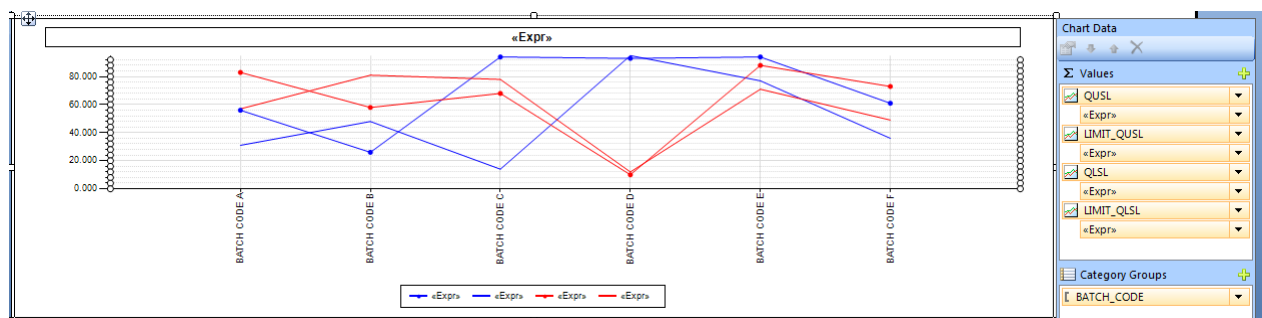
È stato definito un *data set* principale ("DataSetCCData"), per andare a estrarre tutte le informazioni per popolare il *report*, filtrate per i valori definiti nei *prompt* dall'utente (Sorgente D.1).

Gli attributi estratti sono usati, per popolare tutti gli elementi del *report*. Nell'immagine seguente, è possibile notare i valori per definire la prima carta di controllo (Fig. 8.1).



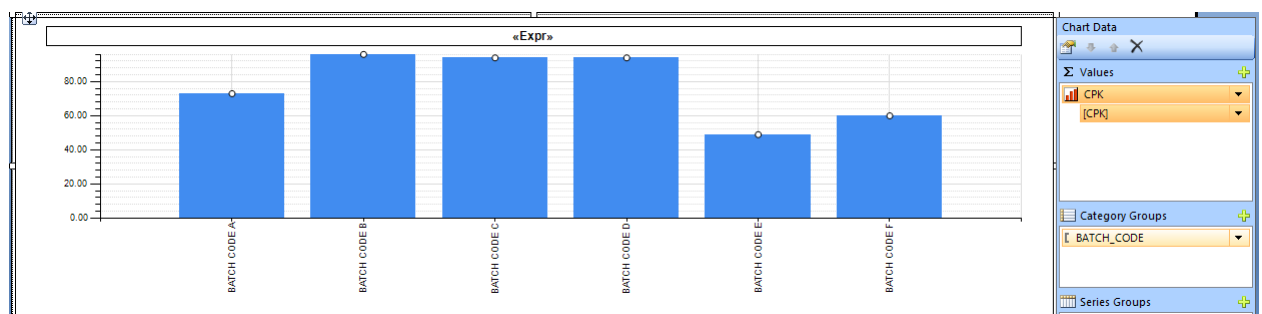
**Figura 8.1:** Definizione della prima carta di controllo

Analogamente, viene definita la seconda carta di controllo (Fig. 8.2).

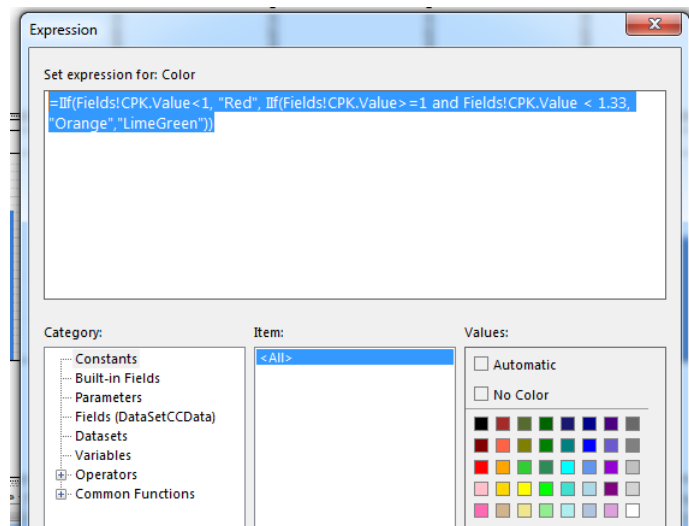


**Figura 8.2:** Definizione della seconda carta di controllo

Per l'istogramma oltre a definire gli attributi per valorizzare il grafico (Fig. 8.3), è stata definita una condizione per il colore che avrà ogni rettangolo in base al valore assunto dal ( $C_{pk}$ ) (Fig. 8.4).

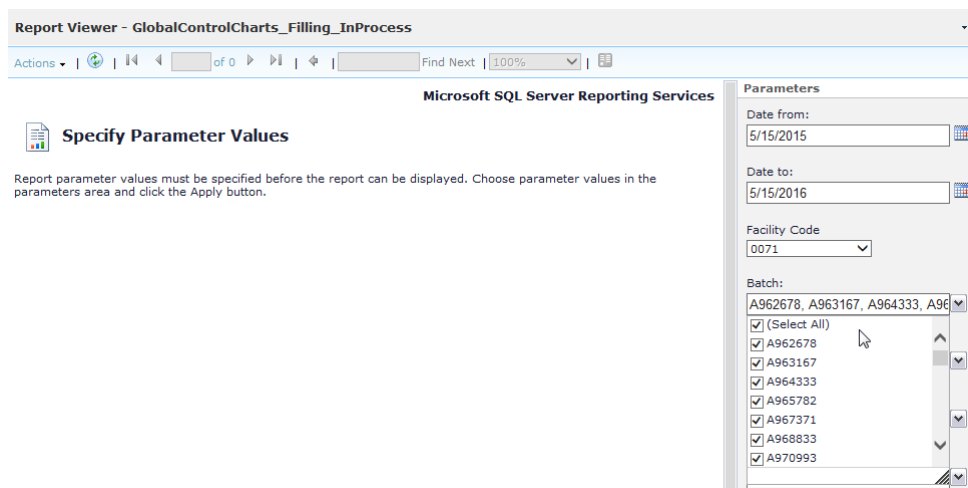


**Figura 8.3:** Definizione dell'istogramma



**Figura 8.4:** Condizione per il colore per l'istogramma

Altri *data set* saranno definiti per andare a valorizzare i *prompt*, come i menù a tendina. Ad esempio per il *prompt* che contiene la lista dei possibili *batch* per la *facility* selezionata precedentemente (Fig. 8.5), è stato definito un altro *data set* (DataSetBatch), per ottenere i valori da visualizzare all'utente (Sorgente D.2).

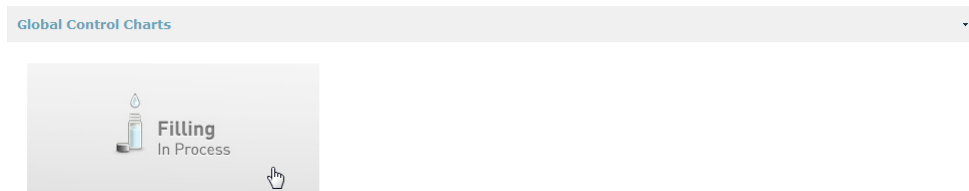


**Figura 8.5:** Menù a tendina con la lista dei possibili *batch*

La *query* selezionerà i valori distinti dei *batch* contenuti nel *data mart* filtrando per la *facility* valorizzata nel *prompt* dall'utente.

## 8.2 Esempi d'uso

Gli utenti abilitati, possono accedere alla soluzione *global filling control charts* tramite un'interfaccia *web* realizzata con "Microsoft SharePoint" (Fig. 8.6).

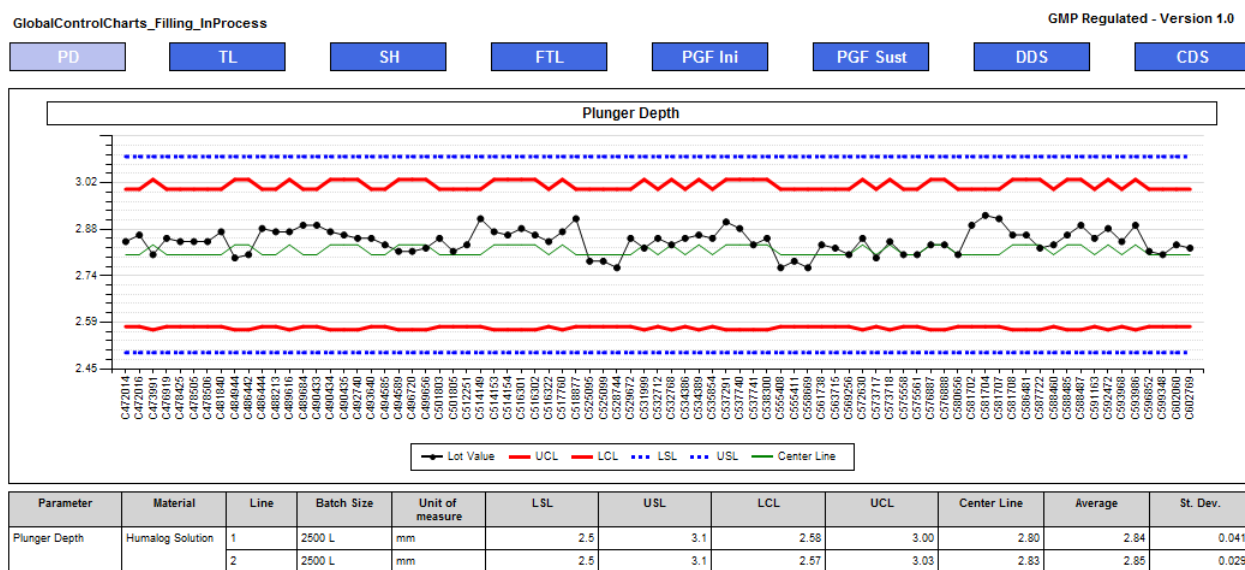


**Figura 8.6:** Interfaccia *web* per le *global filling control charts*

Accedendo l'utente si ritrova sulla destra una lista di *prompt* (a scomparsa) mediante il quale è possibile generare il *report* secondo le sue esigenze di analisi (Fig. 8.7).

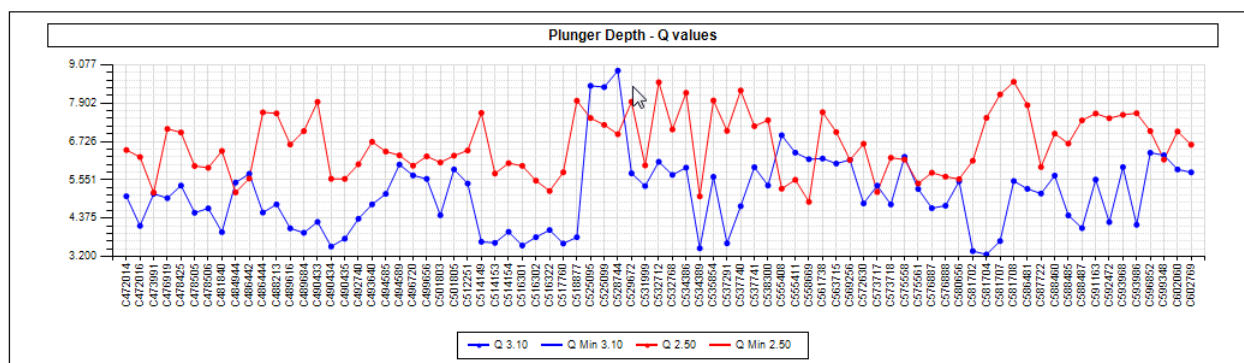
**Figura 8.7:** Finestra principale delle *global filling control charts*

In figura 8.8 è possibile analizzare il parametro *plunger depth* per *batch* di tipo "Humalog Solution".



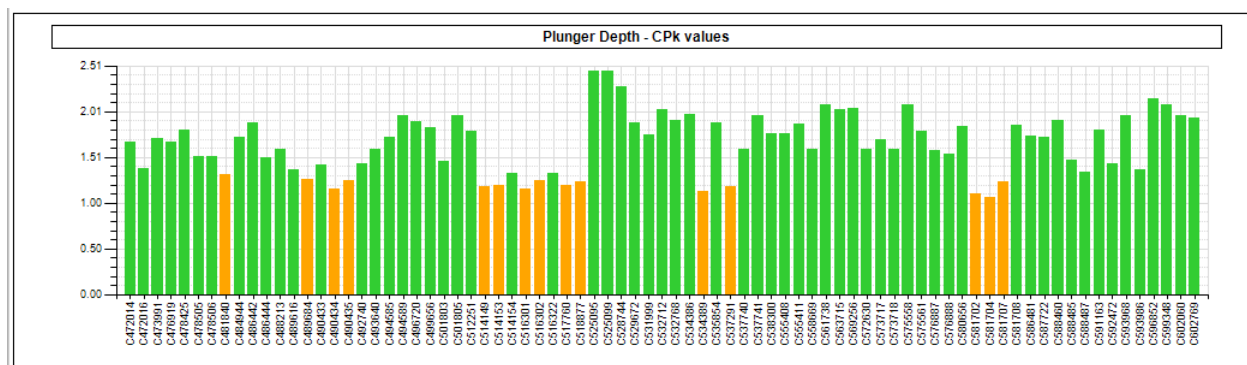
**Figura 8.8:** Esempio di carta di controllo per parametro *plunger depth* per *batch* di tipo Humalog Solution

Scorrendo è possibile visualizzare la seconda carta di controllo contenente per ogni *batch* i valori di LSL e USL (Fig. 8.9).



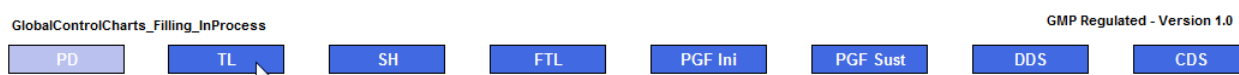
**Figura 8.9:** Esempio di carta di controllo con i valori di LSL e USL per parametro *plunger depth* per *batch* di tipo Humalog Solution

Infine è possibile visualizzare l'istogramma con i valori di ( $C_{pk}$ ) (Fig. 8.10)



**Figura 8.10:** Esempio di istogramma con i valori di  $C_{pk}$  per parametro *plunger depth* per *batch* di tipo Humalog Solution

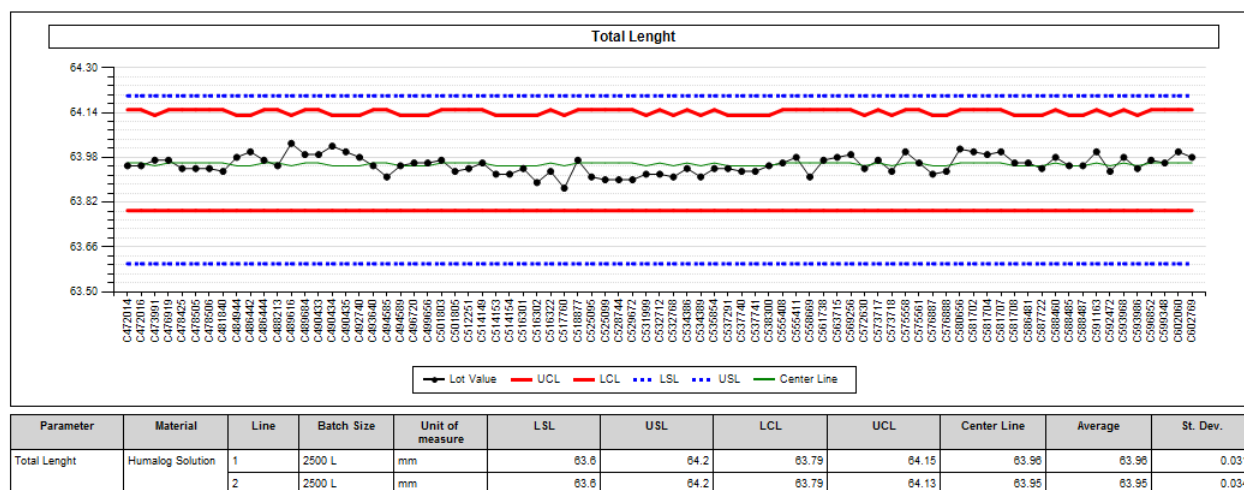
È possibile cambiare parametro di osservazione, semplicemente cliccando nel menù superiore (Fig. 8.11).



**Figura 8.11:** Menù superiore

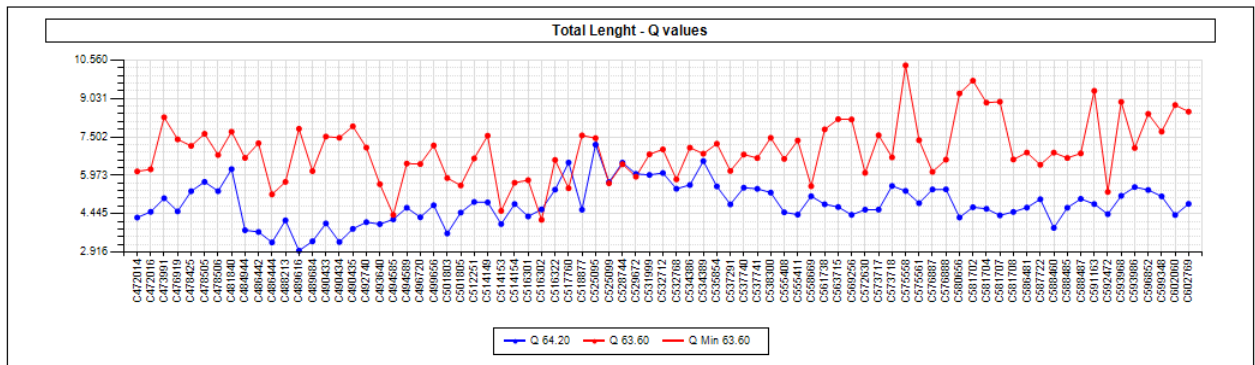
Una volta selezionato un altro parametro di interesse, verrà generato automaticamente il *report* con gli stessi *prompt* impostati precedentemente.

Successivamente si mostrano i grafici ottenuti per il parametro *total length*.

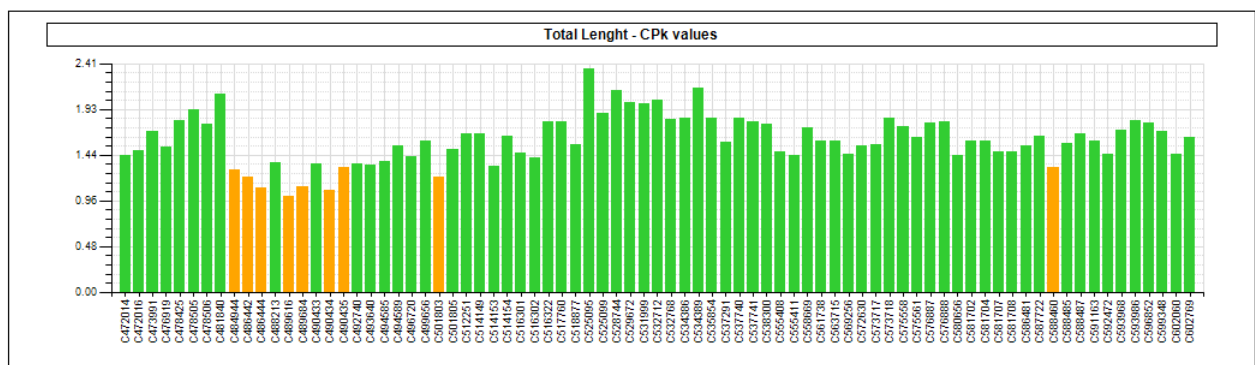


**Figura 8.12:** Esempio di carta di controllo per parametro *total length* per *batch* di tipo Humalog Solution



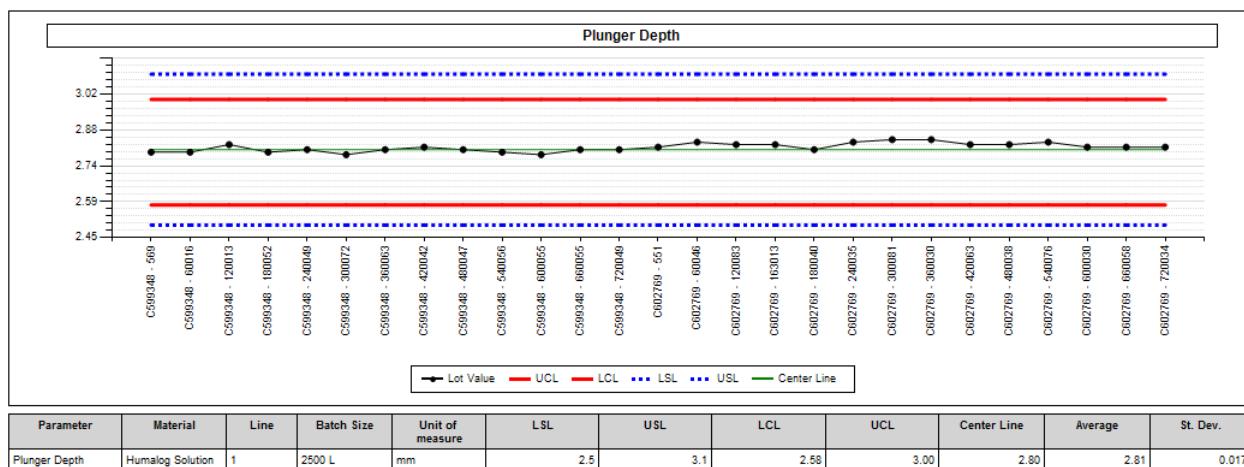


**Figura 8.13:** Esempio di carta di controllo con i valori di LSL e USL per parametro *total length* per batch di tipo Humalog Solution

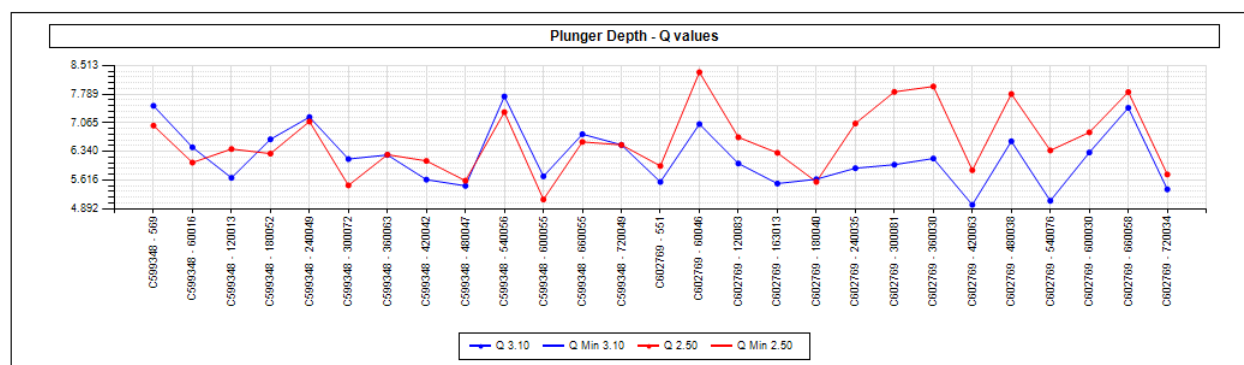


**Figura 8.14:** Esempio di istogramma con i valori di  $C_{pk}$  per parametro *total length* per batch di tipo Humalog Solution

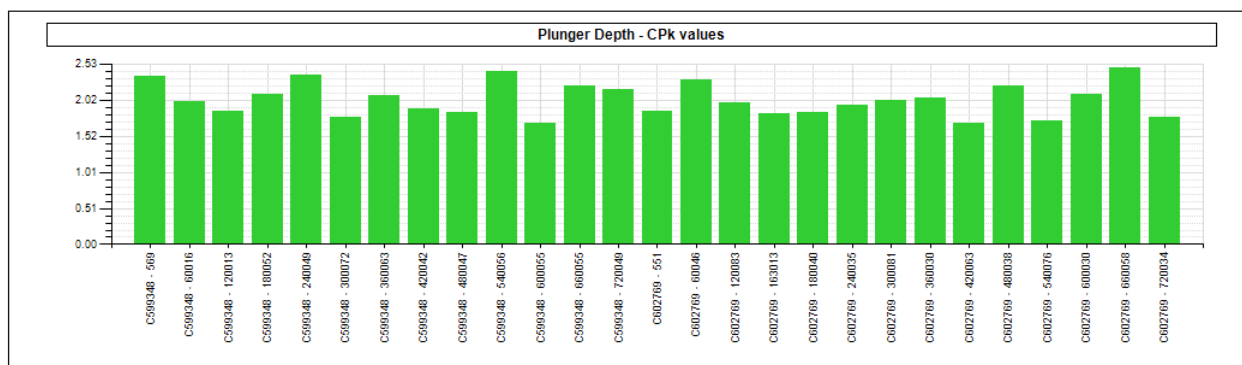
Infine, nelle immagini seguenti è possibile visualizzare il risultato ottenuto attivando la modalità di visualizzazione *sample* su un numero ristretto di *batch*, per rendere il *report* più comprensibile.



**Figura 8.15:** Esempio di carta di controllo per parametro *plunger depth* per *sample* di tipo Humalog Solution



**Figura 8.16:** Esempio di carta di controllo con i valori di LSL e USL per parametro *plunger depth* per *sample* di tipo Humalog Solution



**Figura 8.17:** Esempio di istogramma con i valori di  $C_{pk}$  per parametro *plunger depth* per *sample* di tipo Humalog Solution



# CONCLUSIONI

È stato presentato il lavoro svolto per la progettazione e l'implementazione di un *data warehouse*, portato avanti all'interno di un'attività progettuale dell'azienda SDG Group rivolta ad una importante multinazionale farmaceutica.

L'attività ha permesso di studiare preliminarmente la tematica oggetto del lavoro, come le fasi che compongono il processo produttivo dell'insulina, in particolar modo sul controllo di qualità e sui processi di gestione delle norme di buona fabbricazione, aspetto significativo e di vitale importanza quando al centro dell'attenzione si ha la salute del paziente.

L'obiettivo è stato quello di creare una piattaforma globale, regolamentato dalle norme di buona fabbricazione, dedicata all'analisi dei parametri di interesse misurati durante la fase di riempimento di un lotto di insulina.

Gli sforzi maggiori sono stati sostenuti per definire una corretta trasformazione delle diverse fonti di dato disponibili e la combinazione di queste per ottenere le corrette informazioni da inserire nella struttura dati progettata. Un preciso e strutturato lavoro di ETL è stato infatti implementato e migliorato in modo iterativo per rispettare completamente i requisiti e raffinare il processo.

Il progetto è stato realizzato da un gruppo di due consulenti, per questo motivo ho contribuito a tutte le fasi del progetto, collaborando principalmente alla realizzazione delle procedure ETL e del *report*.

La piattaforma realizzata ha complessivamente soddisfatto le aspettative, ottenendo un riscontro positivo da parte degli utenti finali. Questo ha garantito la continuità collaborativa con il cliente permettendo di definire sviluppi futuri che riguarderanno l'integrazione nel *data mart* di dati provenienti da un altro sito produttivo e la realizzazione della medesima piattaforma per la fase di *sorting*.

Per concludere, SDG Group ha reso questa esperienza lavorativa svolta, particolarmente interessante e stimolante, perché mi ha dato la possibilità di crescere sia sotto il punto di vista umano che quello formativo.

Ho avuto la possibilità di approfondire e mettere in pratica alcune tematiche affrontate durante il corso di studi e di imparare l'utilizzo di strumenti nuovi come "Informatica PowerCenter" e "Report Builder" e approfondire le conoscenze sul DBMS Oracle e il linguaggio PL/SQL.

Si ringrazia il lettore per l'attenzione concessa e coloro che hanno contribuito alla realizzazione di questo lavoro.



# Appendice A

## Estrazione, trasformazione e caricamento

**Listing A.1:** Configurazione della tabella TBSTCNF\_LOAD\_STG

```
1  INSERT
2  INTO
3      TBSTCNF_LOAD_STG
4      (
5          SRC_SYSTEM_CD,
6          SRC_TABLE_NAME_CD,
7          SRC_FACILITY_CD,
8          SRC_INST_CD,
9          HIST_FLG,
10         SRC_DBLINK_CD,
11         STG_SCHEMA_1_LEV,
12         STG_SCHEMA_2_LEV,
13         STG_TABLE_NAME_CD,
14         SRC_FILTER_DESC,
15         JOIN_CONDITION_DESC,
16         DIFF_CONDITION_DESC,
17         LAST_UPDATE_DT,
18         SRC_SCHEMA
19     )
20 VALUES
21     (
22         ' IPC' ,
23         ' IPC_STATS_VIEW' ,
24         ' 0071' ,
25         ' DEF' ,
26         ' N' ,
27         '@DBL_IPC_SES' ,
28         ' GMDM_STG_L1_SES_OWNER' ,
29         ' GMDM_STG_L2_OWNER' ,
30         ' TBSTFT_IPC_STATS_VIEW' ,
31         NULL,
32         ' A.BATCH=B.BATCH AND A.SAMPLING_ID=B.SAMPLING_ID AND A.INSTRUMENT=B.
           INSTRUMENT AND A.MEASURE=B.MEASURE AND A.CALC=B.CALC'
33     ,
34     ' NVL (A.PRODUCT_CODE, ' -1' ) <>nvl (B.PRODUCT_CODE, ' -1
35     ' ) OR
36     NVL (A.IS_BATCH_CLOSE, ' -1' ) <>nvl (B.IS_BATCH_CLOSE, ' -1
37     ' ) OR
```

```

38 NVL(A.CART_PER_PLAN,'-1')<>nvl(B.CART_PER_PLAN,'-1
39 ') OR
40 NVL(A.CART_PRODUCE,-1)<>nvl(B.CART_PRODUCE,-1) OR
41 NVL(A.SAMPLING_TYPE,'
42 -1')<>nvl(B.SAMPLING_TYPE,'-1
43 ') OR
44 NVL(A.VALUE,-1)<>nvl(B.VALUE,-1) OR
45 NVL(A.VALUE_DT,TO_DATE('18530101
46 ','YYYYMMDD'))<>nvl(B.VALUE_DT,TO_DATE('18530101','YYYYMMDD
47 ')) OR
48 NVL(A.LIM_LL,-1)<>nvl(B.LIM_LL,-1) OR
49 NVL(A.LIM_UL,-1)<>nvl(B.LIM_UL,-1) OR
50 NVL(A.NOMINAL,-1)<>nvl(B.NOMINAL,-1)'
51 ,
52 to_date('25-FEB-16 14.38.00','DD-MON-RR HH24.MI.SS'),
53 'XFIPCDISK_OWNER'
54 )

```

---

**Listing A.2:** SQL definito nel *source qualifier* del *mapping* per l'integrazione dell'area del *core*

---

```

1  SELECT
2    BATCH,
3    PRODUCT_CODE,
4    IS_BATCH_CLOSE,
5    SAMPLING_ID,
6    CART_PER_PLAN,
7    CART_PRODUCE,
8    SAMPLING_TYPE,
9    INSTRUMENT,
10   MEASURE,
11   CALC,
12   VALUE,
13   VALUE_DT,
14   LIM_LL,
15   LIM_UL,
16   NOMINAL
17 FROM
18   $$SchemaSTG2L.TBSTFT_IPC_STATS_VIEW
19 WHERE
20   FACILITY_CD      =' $$FCLTY_CD'
21 AND SRC_INST_CD   =' $$SRC_INST_CD'
22 AND RCRD_STS_CD   =' A'
23 AND IS_BATCH_CLOSE = ' Y'
24 AND
25   (
26     (
27       INSTRUMENT = 'ZWICK'
28       AND MEASURE = 'FTL'
29     )
30   OR
31     (
32       INSTRUMENT = 'ZWICK'
33       AND MEASURE LIKE 'PGF%'
34     )
35   OR
36     (
37       INSTRUMENT = 'UPG'

```



---

```

38     )
39     OR
40     (
41         INSTRUMENT = 'DIMCARP'
42     )
43 )
44 AND
45 (
46     UPD_TMSTMP >
47     (
48         SELECT
49             LOAD_CMPRSN_START_TMSTMP
50         FROM
51             $$SCHEMASTG2L.MDM_STG_LOAD_STS
52         WHERE
53             SRC_CD = '$$SRC_CD'
54     )
55 AND UPD_TMSTMP <=
56 (
57     SELECT
58         LOAD_CMPRSN_END_TMSTMP
59     FROM
60         $$SCHEMASTG2L.MDM_STG_LOAD_STS
61     WHERE
62         SRC_CD = '$$SRC_CD'
63 )
64 )

```

---

**Listing A.3:** SQL definito nel *source qualifier* del *mapping* per l'integrazione dell'area dell' *extension*

---

```

1  SELECT
2      A.BATCH_CD,
3      A.MATERIAL_CD,
4      A.PLANNED_SAMPLE_PHASE_CD,
5      A.SAMPLE_ID,
6      A.PARAMETER_CATEGORY_CD,
7      A.PARAMETER_TYPE_CD,
8      A.SAMPLE_TYPE_CD,
9      C.PARAMETER_DESC,
10     B.FILL_LINE_CD,
11     B.FILLING_DT,
12     A.PARAMETER_VAL,
13     A.EFFECTIVE_SAMPLE_NBR,
14     D.BATCH_MFG_DT
15 FROM
16     (
17         SELECT
18             *
19         FROM
20             $$SCHEMAMDM.MDM_INSTRUMENT_MEASURES
21         WHERE
22             FCLTY_CD = '$$FCLTY_CD'
23         AND SRC_INST_CD='$$SRC_INST_CD'
24         AND RCRD_STS_CD='A'
25     )
26     A
27 INNER JOIN

```

```

28  (
29  SELECT
30      MES.BATCH_CD,
31      MES.MATERIAL_CD,
32      MAX(MATLI.FILL_LINE_CD) FILL_LINE_CD,
33      MIN(ORDER_STEP_ACT_SRT_DT ) AS FILLING_DT
34  FROM
35      (
36          SELECT DISTINCT
37              BATCH_CD,
38              MATERIAL_CD
39          FROM
40              $$SCHEMAMDM.MDM_INSTRUMENT_MEASURES
41          WHERE
42              FCLTY_CD      =' $$FCLTY_CD'
43          AND SRC_INST_CD=' $$SRC_INST_CD'
44          AND RCRD_STS_CD=' A'
45      )
46      MES
47  INNER JOIN
48      (
49          SELECT
50              *
51          FROM
52              $$SCHEMAMDM.MDM_MATERIAL_LINES_CNF
53          WHERE
54              FCLTY_CD      =' $$FCLTY_CD'
55          AND SRC_INST_CD=' $$SRC_INST_CD'
56          AND FCC_FLG      =1
57      )
58      MATLI
59  ON
60      MES.MATERIAL_CD=MATLI.MATERIAL_CD
61  INNER JOIN
62      (
63          SELECT
64              *
65          FROM
66              $$SCHEMAMDM.MDM_PROD_ORDER
67          WHERE
68              FCLTY_CD      =' $$FCLTY_CD'
69          AND SRC_INST_CD=' $$SRC_INST_CD'
70          AND RCRD_STS_CD=' A'
71      )
72      PRORD
73  ON
74      PRORD.BATCH_CD      =MES.BATCH_CD
75  AND PRORD.MATERIAL_CD=MES.MATERIAL_CD
76  INNER JOIN
77      (
78          SELECT
79              *
80          FROM
81              $$SCHEMAMDM.MDM_LINE_OPER_PHASE_CNF
82          WHERE
83              FCLTY_CD      =' $$FCLTY_CD'
84          AND SRC_INST_CD      =' $$SRC_INST_CD'

```

```

85         AND WORK_CENTER_DESC='FILLING'
86     )
87     LINEOPER
88     ON
89         LINEOPER.LINE_CD           =MATLI.FILL_LINE_CD
90     AND LINEOPER.OPERATION_CD= PRORD.ORD_STEP_CD
91     GROUP BY
92         MES.BATCH_CD,
93         MES.MATERIAL_CD
94     )
95     B ON A.BATCH_CD =B.BATCH_CD
96     AND A.MATERIAL_CD =B.MATERIAL_CD
97     LEFT OUTER JOIN
98     (
99         SELECT
100             MATERIAL_CD,
101             PARAMETER_CATEGORY_CD,
102             MAX (PARAMETER_DESC) PARAMETER_DESC
103         FROM
104             $$SCHEMAMDM.MDM_PARAMETERS_LIMITS_CNF
105         WHERE
106             FCLTY_CD           =' $$FCLTY_CD'
107         AND SRC_INST_CD       =' $$SRC_INST_CD'
108         AND PARAMETER_TYPE_DESC=' SECONDARY LOOP'
109         GROUP BY
110             MATERIAL_CD,
111             PARAMETER_CATEGORY_CD
112     )
113     C
114     ON
115         A.MATERIAL_CD           =C.MATERIAL_CD
116     AND A.PARAMETER_CATEGORY_CD=C.PARAMETER_CATEGORY_CD
117     LEFT OUTER JOIN
118     (
119         SELECT
120             BATCH_NBR,
121             MAX (BATCH_MFG_DT) BATCH_MFG_DT
122         FROM
123             $$SCHEMAMDM.MDM_BATCH_DIM
124         WHERE
125             FCLTY_CD=' $$FCLTY_CD'
126         GROUP BY
127             BATCH_NBR
128     )
129     D
130     ON
131         D.BATCH_NBR =A.BATCH_CD
132     WHERE
133         A.BATCH_CD NOT IN
134     (
135         SELECT
136             BATCH_CD
137         FROM
138             $$SCHEMAMDMEXT.MDM_CONTROL_CHART_FILLING_DM
139         WHERE
140             FCLTY_CD           =' $$FCLTY_CD'
141         AND SRC_INST_CD=' $$SRC_INST_CD'

```

```

142     AND SRC_CD          =' IPC_SES_HI '
143 )

```

**Listing A.4:** Procedura di *delete* per la tabella del *core*

```

1  CREATE OR REPLACE PROCEDURE DEL_MDM_INSTRUMENT_MEASURES (
2      IN_SRC_FACILITY_CD          IN VARCHAR2,
3      IN_SRC_INST_CD              IN VARCHAR2,
4      IN_SRC_CD                   IN VARCHAR2,
5      IN_PMINTEGRATIONSERVICENAME IN VARCHAR2,
6      IN_PMPREPOSITORYSERVICENAME IN VARCHAR2,
7      IN_PMFOLDERNAME             IN VARCHAR2,
8      IN_PMWORKFLOWNAME           IN VARCHAR2,
9      IN_PMWORKFLOWRUNID          IN VARCHAR2,
10     IN_PMSESSIONNAME             IN VARCHAR2 )
11  IS
12     ERR_MSG                      VARCHAR2 (20000) ;
13     TRNSCTN_TMSTMP_LIMIT DATE;
14  BEGIN
15     SELECT
16         MAX (TABLE_LAST_TRNSCTN_TMSTMP)
17     INTO
18         TRNSCTN_TMSTMP_LIMIT
19     FROM
20         MDM_DELETE_TABLES_CNF
21     WHERE
22         RUNNING_MODE_CD          =' FULL '
23     AND  TABLE_NAME_CD           =' MDM_INSTRUMENT_MEASURES '
24     AND  FCLTY_CD                 =IN_SRC_FACILITY_CD
25     AND  SRC_INST_CD              =IN_SRC_INST_CD
26     AND  SRC_CD                   =IN_SRC_CD;
27     IF TRNSCTN_TMSTMP_LIMIT IS NOT NULL THEN
28         UPDATE
29             MDM_INSTRUMENT_MEASURES
30         SET
31             RCRD_STS_CD  =' D '
32         WHERE
33             NVL (MDM_TRNSCTN_TMSTMP, TO_DATE ('19000101', 'YYYYMMDD')) <
34             TRNSCTN_TMSTMP_LIMIT
35     AND  FCLTY_CD          =IN_SRC_FACILITY_CD
36     AND  SRC_INST_CD       =IN_SRC_INST_CD;
37     INSERT
38     INTO
39         GMDM_GLBL_DISCARD_OWNER.MDM_INSTRUMENT_MEASURES_E
40     (
41         DISCARD_CD,
42         DISCARD_DESC,
43         FCLTY_CD,
44         SRC_INST_CD,
45         RCRD_STS_CD,
46         MDE_FLG,
47         SRC_CD,
48         MDM_LOAD_ID,
49         MDM_TRNSCTN_TMSTMP,
50         MDM_TRNSCTN_TYP_CD,
51         MDM_TRNSCTN_USER_ID,
52         BATCH_CD,

```

```

53         SAMPLE_ID,
54         INSTRUMENT_CD,
55         PARAMETER_CATEGORY_CD,
56         PARAMETER_TYPE_CD
57     )
58     SELECT
59         'DELETE',
60         'DELETE_ATTRIBUTES ( '
61         || IN_SRC_FACILITY_CD
62         || ', '
63         || IN_SRC_INST_CD
64         || ', '
65         || IN_PMINTEGRATIONSERVICENAME
66         || ', '
67         || IN_PMPREPOSITORYSERVICENAME
68         || ', '
69         || IN_PMFOLDERNAME
70         || ', '
71         || IN_PMWORKFLOWNAME
72         || ', '
73         || IN_PMWORKFLOWRUNID
74         || ', '
75         || IN_PMSESSIONNAME
76         || ', '
77         || TO_CHAR (TRNSCTN_TMSTMP_LIMIT, 'YYYY/MM/DD HH24.MI.SS') ,
78         FCLTY_CD,
79         SRC_INST_CD,
80         RCRD_STS_CD,
81         MDE_FLG,
82         SRC_CD,
83         MDM_LOAD_ID,
84         MDM_TRNSCTN_TMSTMP,
85         MDM_TRNSCTN_TYP_CD,
86         MDM_TRNSCTN_USER_ID,
87         BATCH_CD,
88         SAMPLE_ID,
89         INSTRUMENT_CD,
90         PARAMETER_CATEGORY_CD,
91         PARAMETER_TYPE_CD
92     FROM
93         MDM_INSTRUMENT_MEASURES CC
94     WHERE
95         RCRD_STS_CD      ='D'
96     AND FCLTY_CD        =IN_SRC_FACILITY_CD
97     AND SRC_INST_CD     =IN_SRC_INST_CD;
98     DELETE
99     FROM
100         MDM_INSTRUMENT_MEASURES
101     WHERE
102         RCRD_STS_CD      ='D'
103     AND FCLTY_CD        =IN_SRC_FACILITY_CD
104     AND SRC_INST_CD     =IN_SRC_INST_CD;
105     COMMIT;
106     END IF;
107     EXCEPTION
108     WHEN OTHERS THEN
109         ROLLBACK;

```

```

110     RAISE ;
111 END;

```

**Listing A.5:** Procedura di *delete* per la tabella dell'*extension*

```

1  CREATE OR REPLACE PROCEDURE DEL_MDM_CONTROL_CHART_FLLNG_DM (
2      IN_SRC_FACILITY_CD          IN VARCHAR2,
3      IN_SRC_INST_CD              IN VARCHAR2,
4      IN_SRC_CD                   IN VARCHAR2,
5      IN_PMINTEGRATIONSERVICENAME IN VARCHAR2,
6      IN_PMREPOSITORYSERVICENAME  IN VARCHAR2,
7      IN_PMFOLDERNAME             IN VARCHAR2,
8      IN_PMWORKFLOWNAME           IN VARCHAR2,
9      IN_PMWORKFLOWRUNID          IN VARCHAR2,
10     IN_PMSESSIONNAME             IN VARCHAR2 )
11  IS
12     ERR_MSG                      VARCHAR2 (20000) ;
13     TRNSCTN_TMSTMP_LIMIT DATE;
14  BEGIN
15     SELECT
16         MAX (TABLE_LAST_TRNSCTN_TMSTMP)
17     INTO
18         TRNSCTN_TMSTMP_LIMIT
19     FROM
20         GMDM_GLBL_OWNER.MDM_DELETE_TABLES_CNF
21     WHERE
22         RUNNING_MODE_CD          =' FULL'
23     AND  TABLE_NAME_CD          =' MDM_CONTROL_CHART_FILLING_DM'
24     AND  FCLTY_CD                =IN_SRC_FACILITY_CD
25     AND  SRC_INST_CD              =IN_SRC_INST_CD
26     AND  SRC_CD                  =IN_SRC_CD;
27     IF TRNSCTN_TMSTMP_LIMIT IS NOT NULL THEN
28         UPDATE
29             MDM_CONTROL_CHART_FILLING_DM
30         SET
31             RCRD_STS_CD  =' D'
32         WHERE
33             NVL (MDM_TRNSCTN_TMSTMP, TO_DATE ('19000101', 'YYYYMMDD')) <
34             TRNSCTN_TMSTMP_LIMIT
35     AND  FCLTY_CD          =IN_SRC_FACILITY_CD
36     AND  SRC_INST_CD       =IN_SRC_INST_CD
37     AND  SRC_CD            =IN_SRC_CD;
38     INSERT
39     INTO
40         GMDM_GLBL_DISCARD_OWNER.MDM_CONTROL_CHART_FILLING_DM_E
41         (
42             DISCARD_CD,
43             DISCARD_DESC,
44             FCLTY_CD,
45             SRC_INST_CD,
46             RCRD_STS_CD,
47             MDE_FLG,
48             SRC_CD,
49             MDM_LOAD_ID,
50             MDM_TRNSCTN_TMSTMP,
51             MDM_TRNSCTN_TYP_CD,
52             MDM_TRNSCTN_USER_ID,

```

```

53         BATCH_CD,
54         MATERIAL_CD,
55         OPERATION_CD,
56         PHASE_CD,
57         OPERATION_INDEX_VAL,
58         OPERATION_TYPE_DESC,
59         PARAMETER_CD,
60         QUALITY_TYPE_CD
61     )
62     SELECT
63         'DELETE',
64         'DELETE_ATTRIBUTES ( '
65         || IN_SRC_FACILITY_CD
66         || ', '
67         || IN_SRC_INST_CD
68         || ', '
69         || IN_PMINTEGRATIONSERVICENAME
70         || ', '
71         || IN_PMPREPOSITORYSERVICENAME
72         || ', '
73         || IN_PMFOLDERNAME
74         || ', '
75         || IN_PMWORKFLOWNAME
76         || ', '
77         || IN_PMWORKFLOWRUNID
78         || ', '
79         || IN_PMSESSIONNAME
80         || ', '
81         || TO_CHAR (TRNSCTN_TMSTMP_LIMIT, 'YYYY/MM/DD HH24.MI.SS') ,
82         FCLTY_CD,
83         SRC_INST_CD,
84         RCRD_STS_CD,
85         MDE_FLG,
86         SRC_CD,
87         MDM_LOAD_ID,
88         MDM_TRNSCTN_TMSTMP,
89         MDM_TRNSCTN_TYP_CD,
90         MDM_TRNSCTN_USER_ID,
91         BATCH_CD,
92         MATERIAL_CD,
93         OPERATION_CD,
94         PHASE_CD,
95         OPERATION_INDEX_VAL,
96         OPERATION_TYPE_DESC,
97         PARAMETER_CD,
98         QUALITY_TYPE_CD
99     FROM
100         MDM_CONTROL_CHART_FILLING_DM CC
101     WHERE
102         RCRD_STS_CD      =' D'
103     AND FCLTY_CD        =IN_SRC_FACILITY_CD
104     AND SRC_INST_CD     =IN_SRC_INST_CD
105     AND SRC_CD          =IN_SRC_CD;
106     DELETE
107     FROM
108         MDM_CONTROL_CHART_FILLING_DM
109     WHERE

```

```
110      RCRD_STS_CD      ='D'
111      AND FCLTY_CD      =IN_SRC_FACILITY_CD
112      AND SRC_INST_CD   =IN_SRC_INST_CD
113      AND SRC_CD        =IN_SRC_CD;
114      COMMIT;
115      END IF;
116  EXCEPTION
117  WHEN OTHERS THEN
118      ROLLBACK;
119      RAISE ;
120  END;
```

---



## Appendice B

# File dei parametri

**Listing B.1:** *File dei parametri per il workflow wf\_ipc\_gmdm*

---

```
1 [Service:icmdi04-prd]
2 $PMCacheDir=$PMRootDir/gmdm/cache
3 $$PATH_DIR=$PMRootDir/gmdm
4 $PMSessionLogDir=$PMRootDir/gmdm/sesslogs/ipc
5 $PMBadFileDir=$PMRootDir/gmdm/badfiles/ipc
6 $PMWorkflowLogDir=$PMRootDir/gmdm/workflowlogs
7 $PMTempDir=$PMRootDir/gmdm/log/ipc
8 $DBConnectionMDM=GMDM_GLBL_INF_USER
9 $DBConnectionMDMEXT=GMDM_CC_EXT_INF_USER
10 $DBConnectionSTG2L=GMDM_STG_USER_Z2
11 $$FCLTY_CD=0071
12 $$SRC_INST_CD=DEF
13 $$SchemaSTG2L=GMDM_STG_L2_OWNER
14 $$SchemaMDM=GMDM_GLBL_OWNER
15 $$SchemaMDMLOG=GMDM_GLBL_DISCARD_OWNER
16 $$SchemaMDMEXT=GMDM_CC_EXT_OWNER
17 $$INTRFC_TYP_CD=IPC_SES
18 $$APPLCTN_CD=IPC_SES
19 $$SRC_CD=IPC_SES
20 $$DAYOFWEEK_FULL=6
21
22 [s_ins_gmdm_load_sts]
23 $PMSessionLogFile=s_gmdm_ipc_ins_load_sts.log
24 $BadFile1=mdm_load_sts_ins_ipc.bad
25 $BadFile2=gmdm_stg_load_sts_ipc.bad
26
27 [s_upd_gmdm_load_sts]
28 $PMSessionLogFile=s_gmdm_ipc_upd_load_sts.log
29 $BadFile1=mdm_load_sts_upd_ipc.bad
```

---

**Listing B.2:** File dei parametri per il *work flow* wf\_filling\_control\_chart

---

```
1  #[gmdm_ext_project.WF:wf_filling_control_chart_ipcbook_ses]
2  [Service:icmdi04-prd]
3  $PMCacheDir=$PMRootDir/gmdm/cache
4  $PMSessionLogDir=$PMRootDir/gmdm/sesslogs/
5  $PMBadFileDir=$PMRootDir/gmdm/badfiles/
6  $PMWorkflowLogDir=$PMRootDir/gmdm/workflowlogs
7  $PMTempDir=$PMRootDir/gmdm/log/
8  $DBConnectionMDM=GMDM_GLBL_INF_USER
9  $DBConnectionMDMEXT=GMDM_CC_EXT_INF_USER
10 $$SCHEMAMDM=GMDM_GLBL_OWNER
11 $$SCHEMAMDMEXT=GMDM_CC_EXT_OWNER
12 $$SCHEMAMDMLOG=GMDM_GLBL_DISCARD_OWNER
13 $$SCHEMAMDMPROCS=GMDM_CC_EXT_PROCS
14 $$FCLTY_CD=0071
15 $$SRC_CD=IPC_SES
16 $$SRC_INST_CD=DEF
```

---

## Appendice C

# Schedulazione dei *work flow*

**Listing C.1:** Comandi *unix* per la schedulazione dei *work flow*

---

```
1 # begin wf_ipc_gmdm
2
3 # SES
4 00 13 * * 0,1,2,3,4,5,6 export TERM=""; ksh -c '. ~/.kshrc ; pmcmd
    startworkflow -sv ${GMDM_ISERVICE} -d ${GMDM_DOMAIN} -uv GMDM_PMUSER -pv
    GMDM_PMPSWD -f ${GMDM_FOLDER} -rin SES -nowait wf_ipc_gmdm >> $
    GMDM_logdir/scripts/wf_ipc_gmdm.log 2>&1'
5
6 # end wf_ipc_gmdm
7
8 # begin wf_filling_control_chart
9
10 00 14 * * 0,1,2,3,4,5,6 export TERM=""; ksh -c '. ~/.kshrc ; pmcmd
    startworkflow -sv ${GMDM_ISERVICE} -d ${GMDM_DOMAIN} -u gmdm_ext_exec -p
    lilly123 -f gmdm_ext_project -rin SES -nowait wf_filling_control_chart
    >> $GMDM_logdir/scripts/wf_filling_control_chart.log 2>&1'
11
12 # end wf_filling_control_chart
```

---



## Appendice D

# Interfaccia utente e reportistica

Listing D.1: *Data set* DataSetCCData

---

```
1  SELECT
2    BATCH_CD,
3    FCLTY_CD,
4    PARAMETER_TYPE_CD AS PARAMETER_TYPE_STEP_CD,
5    MATERIAL_CD,
6    PARAMETER_CATEGORY_CD,
7    PARAMETER_DESC,
8    PARAMETER_QUANTITY_VAL,
9    PARAMETER_MIN_LIMIT_VAL,
10   PARAMETER_MAX_LIMIT_VAL,
11   PARAMETER_TYPE_CD,
12   NULL AS BATCH_ACT_SRT_DAT,
13   REPORT_CD,
14   MANUFACTURING_DT,
15   BMIN,
16   BMAX,
17   QUSL,
18   QLSL,
19   LIMIT_QUSL,
20   LIMIT_QLSL,
21   Q_LIMITS_DEC_VAL,
22   SIGMA,
23   LOWER_CC_LIMIT,
24   UPPER_CC_LIMIT,
25   PARAMETER_UM_CD,
26   PARAM_DESC_TABLE,
27   LCL_LIMIT_VAL,
28   UCL_LIMIT_VAL,
29   SINGLE_VALUE_DEC_VAL,
30   CCL_DEC_VAL,
31   YTD_CCL_DEC_VAL,
32   AVG_DEC_VAL,
33   CENTER_LINE_DEC_VAL,
34   IHAC_DEC_VAL,
35   DECODE (PARAMETER_CATEGORY_CD, 'FTL' , ROUND ( (PARAMETER_QUANTITY_VAL -
36   PARAMETER_MIN_LIMIT_VAL)                                     / (3
37   *
38   SIGMA), 2), 'PGF Ini' , ROUND ( (PARAMETER_MAX_LIMIT_VAL
39   PARAMETER_QUANTITY_VAL)                                     / (3
40   *
```

```

39  SIGMA), 2), 'PGF Sust' , ROUND ( (PARAMETER_MAX_LIMIT_VAL          -
40  PARAMETER_QUANTITY_VAL)                                           / (3
    *
41  SIGMA), 2), 'CDS' , ROUND ( (PARAMETER_MAX_LIMIT_VAL          -
42  PARAMETER_QUANTITY_VAL)                                           / (3
    *
43  SIGMA), 2), 'DDS' , ROUND ( (PARAMETER_MAX_LIMIT_VAL          -
44  PARAMETER_QUANTITY_VAL)                                           / (3
    *
45  SIGMA), 2), DECODE( SIGN (ROUND ( (PARAMETER_QUANTITY_VAL      -
46  PARAMETER_MIN_LIMIT_VAL)                                           / (3
    *
47  SIGMA), 2)                                                         -
    ROUND ( (
48  PARAMETER_MAX_LIMIT_VAL                                           -
49  PARAMETER_QUANTITY_VAL)                                           / (3
    *
50  SIGMA), 2) ),                                                     -1,
    ROUND
51  ( (PARAMETER_QUANTITY_VAL                                           -
52  PARAMETER_MIN_LIMIT_VAL)                                           / (3
    *
53  SIGMA), 2), ROUND ( (PARAMETER_MAX_LIMIT_VAL                    -
54  PARAMETER_QUANTITY_VAL)                                           / (3
    *
55  SIGMA), 2) ) ) CPK,
56  AVG_CHART_FLG,
57  AVG_TABLE_FLG,
58  LIHAC_FLG,
59  UIHAC_FLG,
60  LCL_FLG,
61  UCL_FLG,
62  CENTER_LINE_FLG,
63  STD_DEV_FLG,
64  YTD_LCL_FLG,
65  YTD_UCL_FLG,
66  CPK_FLG,
67  FILLING_DT,
68  OTHER_DEC_VAL,
69  MATERIAL_SHORT_DESC,
70  BATCH_SIZE_CD,
71  FILL_LINE_CD,
72  CENTER_LINE_VAL,
73  (
74  CASE
75      WHEN NVL (LIMIT_QUSL, 99999) < NVL (QUSL, 99999)
76      AND NVL (LIMIT_QUSL, 99999) < NVL (QLSL, 99999)
77      AND NVL (LIMIT_QUSL, 99999) < NVL (LIMIT_QLSL, 99999)
78      THEN LIMIT_QUSL
79      WHEN NVL (LIMIT_QLSL, 99999) < NVL (QUSL, 99999)
80      AND NVL (LIMIT_QLSL, 99999) < NVL (QLSL, 99999)
81      THEN LIMIT_QLSL
82      WHEN NVL (QUSL, 99999) < NVL (QLSL, 99999)
83      THEN QUSL
84      ELSE QLSL
85  END) QL_GRAPH_MIN,
86  (

```

```

87  CASE
88      WHEN NVL(LIMIT_QUSL,-99999)>NVL(QUSL,-99999)
89      AND NVL(LIMIT_QUSL,-99999)>NVL(QLSL,-99999)
90      AND NVL(LIMIT_QUSL,-99999)>NVL(LIMIT_QLSL,-99999)
91      THEN LIMIT_QUSL
92      WHEN NVL(LIMIT_QLSL,-99999)>NVL(QUSL,-99999)
93      AND NVL(LIMIT_QLSL,-99999)>NVL(QLSL,-99999)
94      THEN LIMIT_QLSL
95      WHEN NVL(QUSL,-99999)>NVL(QLSL,-99999)
96      THEN QUSL
97      ELSE QLSL
98  END) QL_GRAPH_MAX
99  FROM
100  (
101      SELECT DISTINCT
102          Y.BATCH_CD
103          ||DECODE(:SHOW_SAMPLE,'SAMPLE',' - '
104          ||Y.STAT_SAMPLE_SIZE_NBR,'') AS BATCH_CD,
105          Y.FCLTY_CD,
106          Y.PARAMETER_TYPE_CD,
107          Y.MATERIAL_CD,
108          Y.PARAMETER_CATEGORY_CD,
109          Y.PARAMETER_DESC,
110          Y.PARAMETER_QUANTITY_VAL,
111          PLIM.PARAMETER_MIN_LIMIT_VAL,
112          PLIM.PARAMETER_MAX_LIMIT_VAL,
113          Y.REPORT_CD,
114          Y.MANUFACTURING_DT,
115          PLIM.LCL_LIMIT_VAL LOWER_CC_LIMIT,
116          PLIM.UCL_LIMIT_VAL UPPER_CC_LIMIT,
117          PLIM.PARAMETER_UM_CD,
118          PLIM.PARAMETER_DESC PARAM_DESC_TABLE,
119          PLIM.LCL_LIMIT_VAL,
120          PLIM.UCL_LIMIT_VAL,
121          PLIM.SINGLE_VALUE_DEC_VAL,
122          PLIM.CCL_DEC_VAL,
123          PLIM.YTD_CCL_DEC_VAL,
124          PLIM.AVG_DEC_VAL,
125          PLIM.CENTER_LINE_DEC_VAL,
126          PLIM.IHAC_DEC_VAL,
127          PLIM.CENTER_LINE_VAL,
128          AVG_CHART_FLG,
129          AVG_TABLE_FLG,
130          LIHAC_FLG,
131          UIHAC_FLG,
132          LCL_FLG,
133          UCL_FLG,
134          CENTER_LINE_FLG,
135          STD_DEV_FLG,
136          YTD_LCL_FLG,
137          YTD_UCL_FLG,
138          CPK_FLG,
139          FILLING_DT,
140          OTHER_DEC_VAL,
141          BATCH_SIZE_CD,
142          FILLING_LINE_CD          AS FILL_LINE_CD,
143          Y.BATCH_CD              AS BATCH_CD_ORD,

```

```

144 Y.STAT_SAMPLE_SIZE_NBR AS STAT_SAMPLE_SIZE_NBR_ORD,
145 MATERIAL_SHORT_DESC,
146 (
147     SELECT DISTINCT
148         X.PARAMETER_QUANTITY_VAL
149     FROM
150         GMDM_CC_EXT_OWNER.MDM_CONTROL_CHART_FILLING_DM X
151     WHERE
152         X.REPORT_CD = Y.REPORT_CD
153     AND X.PARAMETER_CATEGORY_CD = Y.PARAMETER_CATEGORY_CD
154     AND X.BATCH_CD = Y.BATCH_CD
155     AND X.MATERIAL_CD = Y.MATERIAL_CD
156     AND X.FCLTY_CD = Y.FCLTY_CD
157     AND X.SRC_INST_CD = Y.SRC_INST_CD
158     AND X.OPERATION_INDEX_VAL = Y.OPERATION_INDEX_VAL
159     AND X.PARAMETER_TYPE_CD = 'Min'
160 )
161 BMIN,
162 (
163     SELECT DISTINCT
164         X.PARAMETER_QUANTITY_VAL
165     FROM
166         GMDM_CC_EXT_OWNER.MDM_CONTROL_CHART_FILLING_DM X
167     WHERE
168         X.REPORT_CD = Y.REPORT_CD
169     AND X.PARAMETER_CATEGORY_CD = Y.PARAMETER_CATEGORY_CD
170     AND X.BATCH_CD = Y.BATCH_CD
171     AND X.MATERIAL_CD = Y.MATERIAL_CD
172     AND X.FCLTY_CD = Y.FCLTY_CD
173     AND X.SRC_INST_CD = Y.SRC_INST_CD
174     AND X.OPERATION_INDEX_VAL = Y.OPERATION_INDEX_VAL
175     AND X.PARAMETER_TYPE_CD = 'Max'
176 )
177 BMAX,
178 (
179     SELECT DISTINCT
180         X.PARAMETER_QUANTITY_VAL
181     FROM
182         GMDM_CC_EXT_OWNER.MDM_CONTROL_CHART_FILLING_DM X
183     WHERE
184         X.REPORT_CD = Y.REPORT_CD
185     AND X.PARAMETER_CATEGORY_CD = Y.PARAMETER_CATEGORY_CD
186     AND X.BATCH_CD = Y.BATCH_CD
187     AND X.MATERIAL_CD = Y.MATERIAL_CD
188     AND X.FCLTY_CD = Y.FCLTY_CD
189     AND X.SRC_INST_CD = Y.SRC_INST_CD
190     AND X.OPERATION_INDEX_VAL = Y.OPERATION_INDEX_VAL
191     AND X.PARAMETER_TYPE_CD = 'USL'
192 )
193 QUSL,
194 (
195     SELECT DISTINCT
196         X.PARAMETER_QUANTITY_VAL
197     FROM
198         GMDM_CC_EXT_OWNER.MDM_CONTROL_CHART_FILLING_DM X
199     WHERE
200         X.REPORT_CD = Y.REPORT_CD

```



```

201      AND X.PARAMETER_CATEGORY_CD = Y.PARAMETER_CATEGORY_CD
202      AND X.BATCH_CD              = Y.BATCH_CD
203      AND X.MATERIAL_CD           = Y.MATERIAL_CD
204      AND X.FCLTY_CD              = Y.FCLTY_CD
205      AND X.SRC_INST_CD           = Y.SRC_INST_CD
206      AND X.OPERATION_INDEX_VAL   = Y.OPERATION_INDEX_VAL
207      AND X.PARAMETER_TYPE_CD     = 'LSL'
208  )
209  QLSL,
210  (
211      SELECT DISTINCT
212          X.PARAMETER_QUANTITY_VAL
213      FROM
214          GMDM_CC_EXT_OWNER.MDM_CONTROL_CHART_FILLING_DM X
215      WHERE
216          X.REPORT_CD              = Y.REPORT_CD
217      AND X.PARAMETER_CATEGORY_CD = Y.PARAMETER_CATEGORY_CD
218      AND X.BATCH_CD              = Y.BATCH_CD
219      AND X.MATERIAL_CD           = Y.MATERIAL_CD
220      AND X.FCLTY_CD              = Y.FCLTY_CD
221      AND X.SRC_INST_CD           = Y.SRC_INST_CD
222      AND X.OPERATION_INDEX_VAL   = Y.OPERATION_INDEX_VAL
223      AND X.PARAMETER_TYPE_CD     = 'Sigma'
224  )
225  SIGMA,
226  (
227      SELECT
228          MAX(OTHER_MIN_LIMIT_VAL)
229      FROM
230          GMDM_GLBL_OWNER.MDM_PARAMETERS_LIMITS_CNF PARLIMS
231      WHERE
232          PARAMETER_CATEGORY_CD LIKE DECODE (:MEASURE, 'PD', :MEASURE
233          || '%Q%Min%3.1', 'noose')
234      AND Y.FCLTY_CD              = PARLIMS.FCLTY_CD
235      AND Y.SRC_INST_CD           = PARLIMS.SRC_INST_CD
236      AND LIMIT_START_DT <= DECODE (:FILTER_DATE, 'Manufacturing',
237          MANUFACTURING_DT, FILLING_DT)
238      AND LIMIT_expir_DT >= DECODE (:FILTER_DATE, 'Manufacturing',
239          MANUFACTURING_DT, FILLING_DT)
240  )
241  LIMIT_QUSL,
242  (
243      SELECT
244          MAX(OTHER_MIN_LIMIT_VAL)
245      FROM
246          GMDM_GLBL_OWNER.MDM_PARAMETERS_LIMITS_CNF PARLIMS
247      WHERE
248          PARAMETER_CATEGORY_CD LIKE DECODE (:MEASURE, 'PD', :MEASURE
249          || '%Q%Min%2.5', :MEASURE
250          || '%Q%Min')
251      AND Y.FCLTY_CD              = PARLIMS.FCLTY_CD
252      AND Y.SRC_INST_CD           = PARLIMS.SRC_INST_CD
253      AND LIMIT_START_DT <= DECODE (:FILTER_DATE, 'Manufacturing',
254          MANUFACTURING_DT, FILLING_DT)
255      AND LIMIT_expir_DT >= DECODE (:FILTER_DATE, 'Manufacturing',
256          MANUFACTURING_DT, FILLING_DT)
257  )

```

```

258     LIMIT_QLSL,
259     NVL (
260     (
261         SELECT
262             MAX(OTHER_DEC_VAL)
263         FROM
264             GMDM_GLBL_OWNER.MDM_PARAMETERS_LIMITS_CNF PARLIMS
265         WHERE
266             PARAMETER_CATEGORY_CD LIKE DECODE (:MEASURE, 'PD', :MEASURE
267             || '%Q%Min%2.5', :MEASURE
268             || '%Q%Min')
269         AND Y.FCLTY_CD      = PARLIMS.FCLTY_CD
270         AND Y.SRC_INST_CD   = PARLIMS.SRC_INST_CD
271         AND LIMIT_START_DT <= DECODE (:FILTER_DATE, 'Manufacturing',
272             MANUFACTURING_DT, FILLING_DT)
273         AND LIMIT_expir_DT >= DECODE (:FILTER_DATE, 'Manufacturing',
274             MANUFACTURING_DT, FILLING_DT)
275     )
276     , 3) Q_LIMITS_DEC_VAL
277 FROM
278     (
279         SELECT
280             *
281         FROM
282             GMDM_CC_EXT_OWNER.MDM_CONTROL_CHART_FILLING_DM
283         WHERE
284             FCLTY_CD=:FCLTY_CD
285     )
286     Y
287 INNER JOIN
288     (
289         SELECT
290             *
291         FROM
292             GMDM_GLBL_OWNER.MDM_PARAMETERS_LIMITS_CNF
293         WHERE
294             FCLTY_CD=:FCLTY_CD
295     )
296     PLIM
297 ON
298     (
299         Y.PARAMETER_CATEGORY_CD = PLIM.PARAMETER_CATEGORY_CD
300     AND
301         (
302             (
303                 Y.MATERIAL_CD = (
304                     CASE
305                         WHEN MANUFACTURING_DT <= TO_DATE('31-12-2013','DD-MM-YYYY')
306                         AND Y.FCLTY_CD      ='0071'
307                         THEN PLIM.MATERIAL_MATCH_CD
308                         ELSE PLIM.MATERIAL_CD
309                     END )
310             ) --ECCEZIONE PER PRECEDENTE GESTIONE BATCH
311         OR PLIM.MATERIAL_CD = 'ALL'
312     )
313     AND
314     (

```

```

315         PLIM.LIMIT_START_DT    <= Y.FILLING_DT
316     AND PLIM.LIMIT_EXPIR_DT >= Y.FILLING_DT
317 )
318 AND Y.PARAMETER_TYPE_CD IN ('Avg')
319 AND Y.SRC_INST_CD      =PLIM.SRC_INST_CD
320 )
321 INNER JOIN
322 (
323     SELECT
324         *
325     FROM
326         GMDM_GLBL_OWNER.MDM_MATERIAL_BATCH_TYPE_AS_CNF
327     WHERE
328         FCLTY_CD =:FCLTY_CD
329     AND FCC_FLG=1
330 )
331 C
332 ON
333     C.MATERIAL_CD          = Y.MATERIAL_CD
334 AND C.SRC_INST_CD        =Y.SRC_INST_CD
335 AND Y.PARAMETER_CATEGORY_CD = :MEASURE
336 AND Y.MATERIAL_CD        IN (:MATERIAL)
337 AND BATCH_CD             IN (:BATCH)
338 AND
339 (
340     (
341         :SHOW_SAMPLE          =' BATCH'
342     AND Y.OPERATION_INDEX_VAL =0
343     )
344 OR
345     (
346         :SHOW_SAMPLE          =' SAMPLE'
347     AND Y.OPERATION_INDEX_VAL >0
348     )
349 )
350 AND MATERIAL_SHORT_DESC IN (:MATERIALSHORTDESC)
351 AND BATCH_SIZE_CD      IN (:BATCHSIZE)
352 AND FILLING_LINE_CD    IN (:FILLLINE)
353 AND
354 (
355     DECODE(:FILTER_DATE, 'Manufacturing', MANUFACTURING_DT, FILLING_DT)
356     BETWEEN NVL(:MNFCT_DATE_FROM, TO_DATE('01/01/1900', 'DD/MM/RRRR'))
357             AND
358             NVL(:MNFCT_DATE_TO, TO_DATE('31/12/9999', 'DD/MM/RRRR'))
359 OR
360 (
361     DECODE(:FILTER_DATE, 'Manufacturing', MANUFACTURING_DT,
362             FILLING_DT)
363     IS NULL
364 )
365 )
366 ORDER BY
367 CASE
368     WHEN :SORT = 'Batch'
369     THEN Y.BATCH_CD
370 END,
371 CASE

```

---

```

370         WHEN :SORT = 'Batch'
371         THEN Y.STAT_SAMPLE_SIZE_NBR
372     END,
373     CASE
374         WHEN :SORT = 'Date'
375         THEN DECODE(:FILTER_DATE, 'Manufacturing', MANUFACTURING_DT,
376                     FILLING_DT
377         )
378     END,
379     CASE
380         WHEN :SORT = 'Date'
381         THEN Y.BATCH_CD
382     END,
383     CASE
384         WHEN :SORT = 'Date'
385         THEN Y.STAT_SAMPLE_SIZE_NBR
386     END
387 )
    CC

```

---

#### Listing D.2: *Data set DataSetBatch*

---

```

1  SELECT DISTINCT
2      BATCH_CD
3  FROM
4      GMDM_CC_EXT_OWNER.MDM_CONTROL_CHART_FILLING_DM
5  WHERE
6      REPORT_CD = 'SECONDARY LOOP'
7  AND FCLTY_CD=:Fclty_cd
8  ORDER BY
9      BATCH_CD

```

---

# Bibliografia

- [Albano,2014] Antonio Albano, *Decision support databases essential*, Appunti delle lezioni del corso di Sistemi informatici direzionali. (Citato alle pagine 1, 17, 18, 25, 27 e 29)
- [Coccarda,2011] Raoul Coccarda, *Manuale di statistica*, Statistica descrittiva, inferenziale e calcolo delle probabilità. (Citato a pagina 7)
- [Gartner,2015] Gartner, *Magic quadrant for data integration tools*, <https://www.gartner.com/doc/3102119/magic-quadrant-data-integration-tools> (Citato a pagina 39)
- [Gartner,2015] Gartner, *Magic quadrant for operational database management systems*, <https://www.gartner.com/doc/reprints?id=1-2PMFPEN&ct=151013&st=sb> (Citato a pagina 38)
- [Informatica,2013] Informatica Corporation, *Informatica PowerCenter express*, User guide. (Citato alle pagine 2 e 39)
- [Lanati,2010] Antonella Lanati, *Qualità in biotech e pharma*, Gestione manageriale dei processi dalla ricerca ai suoi prodotti. (Citato alle pagine 2 e 5)
- [Lanzara,2012] Riccardo Lanzara, *La rosa e il sensore*, tradizione e innovazione nell'impresa dinamica. (Citato alle pagine 2 e 4)
- [Miolo Vitali,2000] Paola Miolo Vitali, *Corso di economia aziendale*, Decisioni, processi decisionali e misurazioni. (Citato alle pagine 2 e 3)
- [OSI Software,2016] *Profilo aziendale di OSI Software*, <http://www.osisoft.com>. (Citato a pagina 32)
- [Rockwell Automation,2016] *Profilo aziendale di Rockwell Automation*, <http://www.rockwellautomation.com/>. (Citato a pagina 33)
- [SDG Group,2016] *Profilo aziendale di SDG Group*, <http://www.sdggroup.com>. (Citato alle pagine 2 e 12)
- [Seidenader,2016] *Profilo aziendale di Seidenader*, <http://www.seidenader.de/inspection/automatic-inspection>. (Citato a pagina 15)
- [UNI EN ISO 8402,1995] UNI EN ISO 8402, *Gestione per la qualità ed assicurazione della qualità*, Termini e definizioni. (Citato alle pagine 2 e 5)